



# Travel Insurance

Analysis & Machine Learning Model

**Final Presentation**

**Group 6 – Batch 18**



# Meet our Teams



Alya  
Aliski



Muh. Ariq  
Arfina



Jordy  
Ichwansyah



Ni Gusti Made  
Yogeta Yushanti



Muh. Tidar  
Wirayudha



Lewis  
Jonathan



M Fauzan



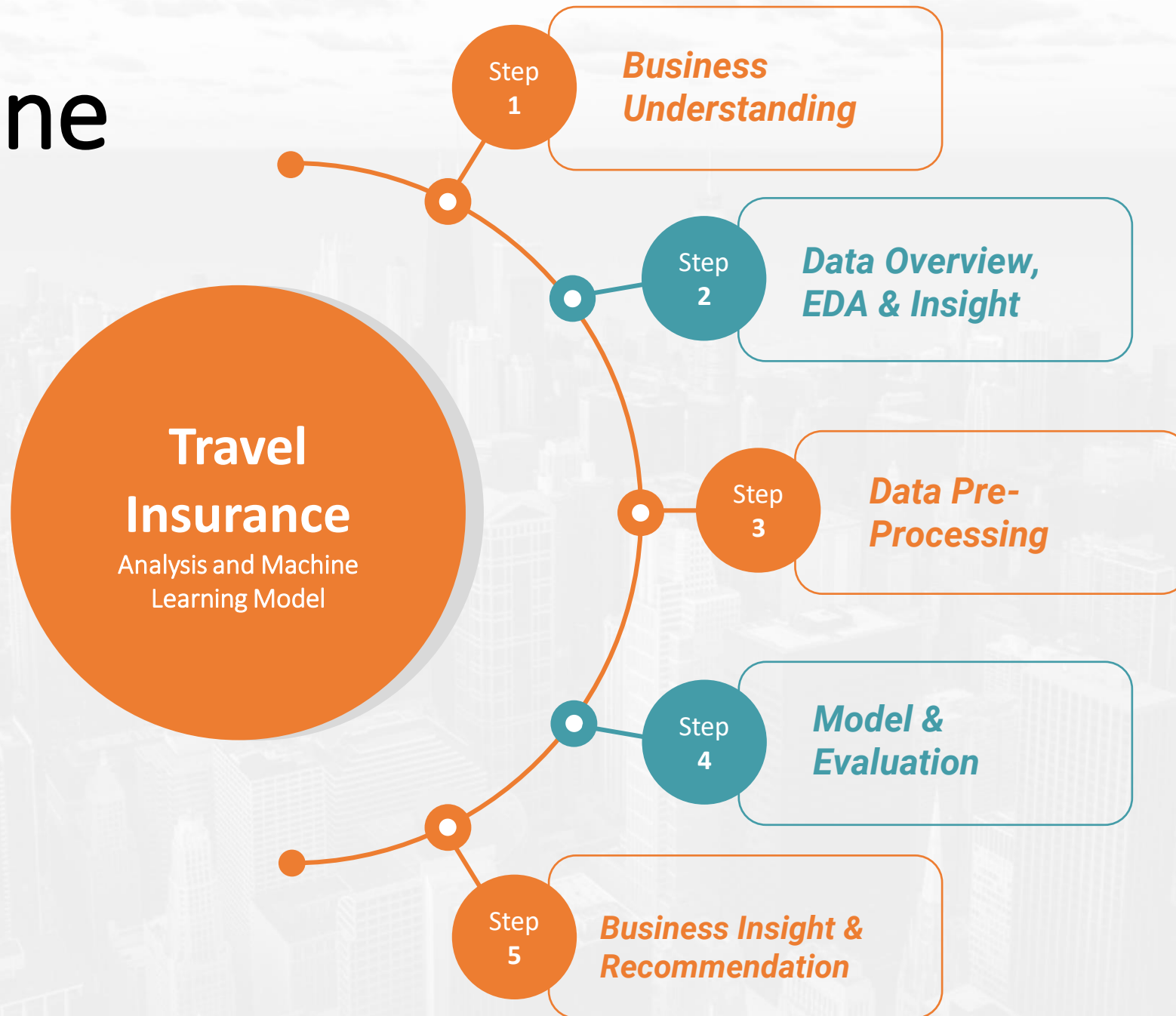
# Who Are We?

**HemaPython** adalah konsultan Data Scientist yang siap memberikan rekomendasi dan aksi untuk menyelesaikan masalah pada Perusahaan Start-up





# Outline



# *Business Understanding*



# Who Are The Client and What Case?

**TravelBuddy** merupakan sebuah perusahaan Start-up baru yang bergerak di bidang Tour and Travel.

Mereka mempunyai sebuah produk baru yaitu **Travel Insurance**, dengan Benefit sebagai berikut:

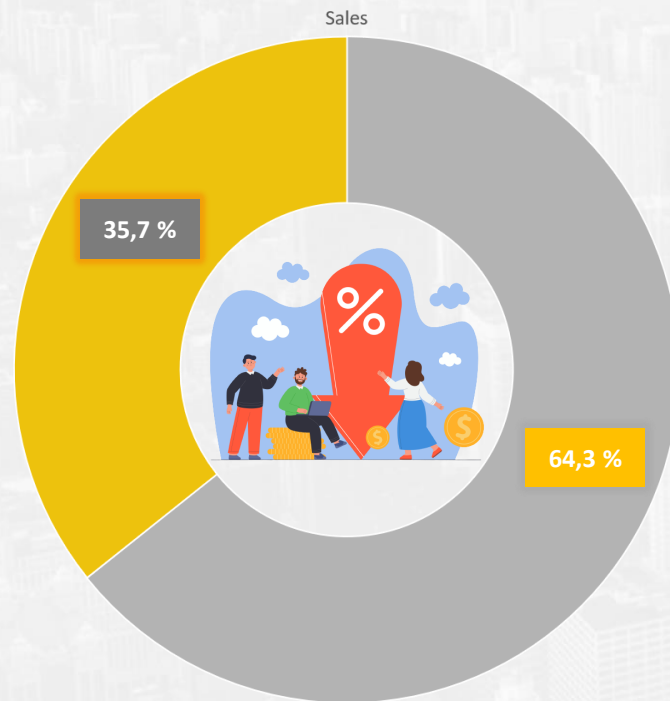
- Death due to Covid-19
- Hospital cash plan
- Income Loss Protection

Tim marketing TravelBuddy **belum melakukan targeted marketing sejak tahun 2020 dan channel yang masih terbatas** untuk menawarkan produk Travel Insurance. Sehingga di tahun 2021 mereka **mengalami kerugian** karena **biaya marketing terlalu besar**, sehingga menyebabkan **penjualan produk Travel Insurance tidak signifikan**.



# Let's Check The Data

## TravelBuddy Travel Insurance Sales



- Di Tahun 2020, TravelBuddy memiliki total pelanggan sebanyak **1987**.
- Dari penjualan yang dilakukan selama 1 tahun, hanya **35,7%** yang membeli produk Travel Insurance!

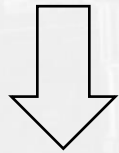
Marketing yang tidak tertarget bisa menjadi salah satu alasan mengapa penjualan tidak signifikan. Jika marketing tidak tertarget, **maka pengeluaran untuk biaya marketing akan membengkak.**

**Jika di Asumsikan :**  
Marketing Cost per customer  
: Rp 100.000,- /Cust

**Marketing Cost 2020:**  
**Rp 198.700.000,-**

# Problem Statement

**Marketing tidak tertarget**



**Marketing cost membengkak**





# Goals, Objective, Business Metrics



- **Goals :**
  - **Memprediksi** customer yang berpotensi **membeli Travel Insurance** ( $> 75\%$ )
  - Mengurangi **Marketing Cost**
- **Objective :**
  - Membuat sebuah **machine learning model** yang dapat **memprediksi customer yang tertarik** untuk **membeli Travel Insurance**
  - Memberikan **insight** dan **rekomendasi aksi** yang dapat membantu product team dan business
- **Business Metrics**  
Customer Acquisition Cost

# *Data Overview, EDA & Insight*



# Data Overview

## Travel Insurance Prediction Data

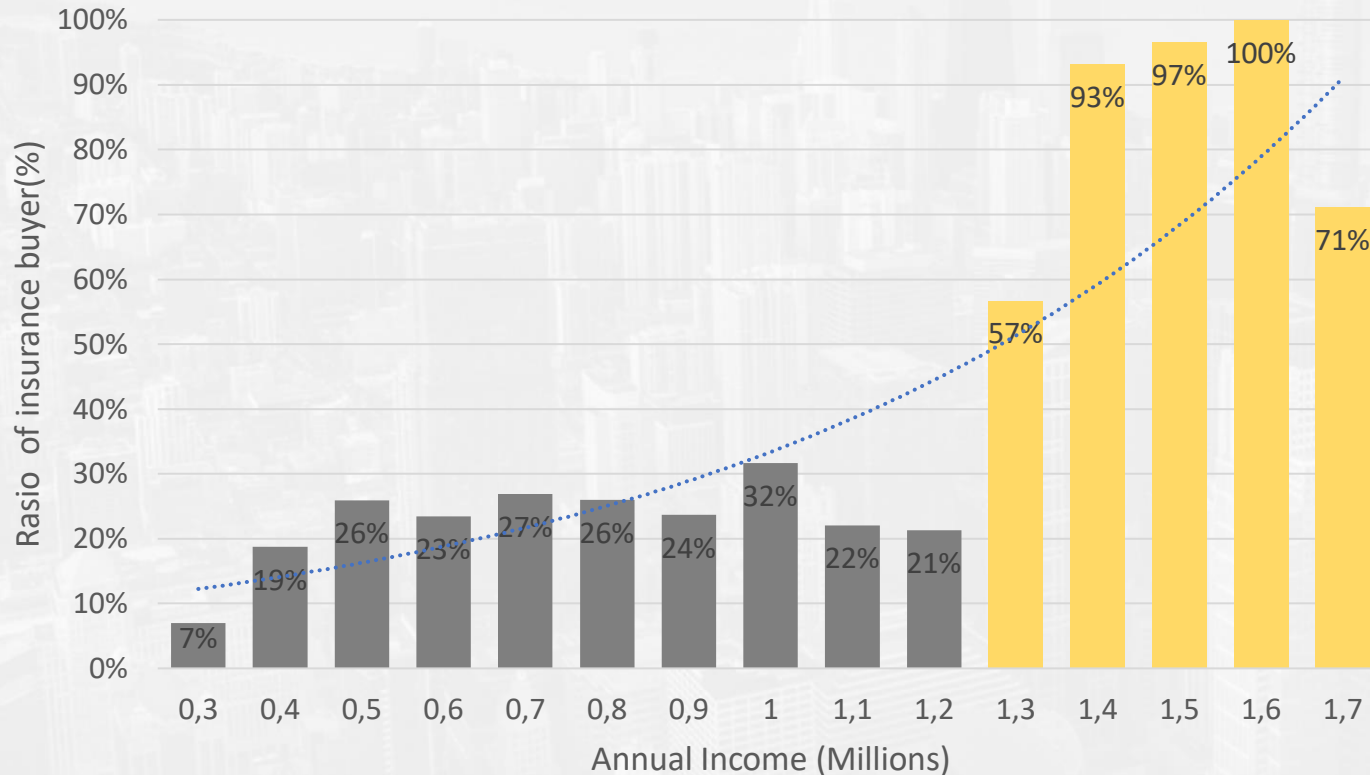
| Age                 | Age Of The Customer   |
|---------------------|---|
| Employment Type     | The Sector In Which Customer Is Employed  |
| GraduateOrNot       | Whether The Customer Is College Graduate Or Not   |
| AnnualIncome        | The Yearly Income Of The Customer In <b>Indian Rupees</b> [Rounded To Nearest 50 Thousand Rupees]                               |
| FamilyMembers       | Number Of Members In Customer's Family  |
| ChronicDisease      | Whether The Customer Suffers From Any Major Disease Or Conditions Like Diabetes/High BP or Asthama,etc.                         |
| FrequentFlyer       | Derived Data Based On Customer's History Of Booking Air Tickets On Atleast 4 Different Instances In The Last 2 Years[2017-2019] |
| EverTravelledAbroad | Has The Customer Ever Travelled To A Foreign Country[Not Necessarily Using The Company's Services]                              |
| TravellInsurance    | Did The Customer Buy Travel Insurance Package During Introductory Offering Held In The Year 2019.                               |

Total Feature : 9

Target Label : **TravellInsurance**



# Exploratory Data Analysis (EDA)

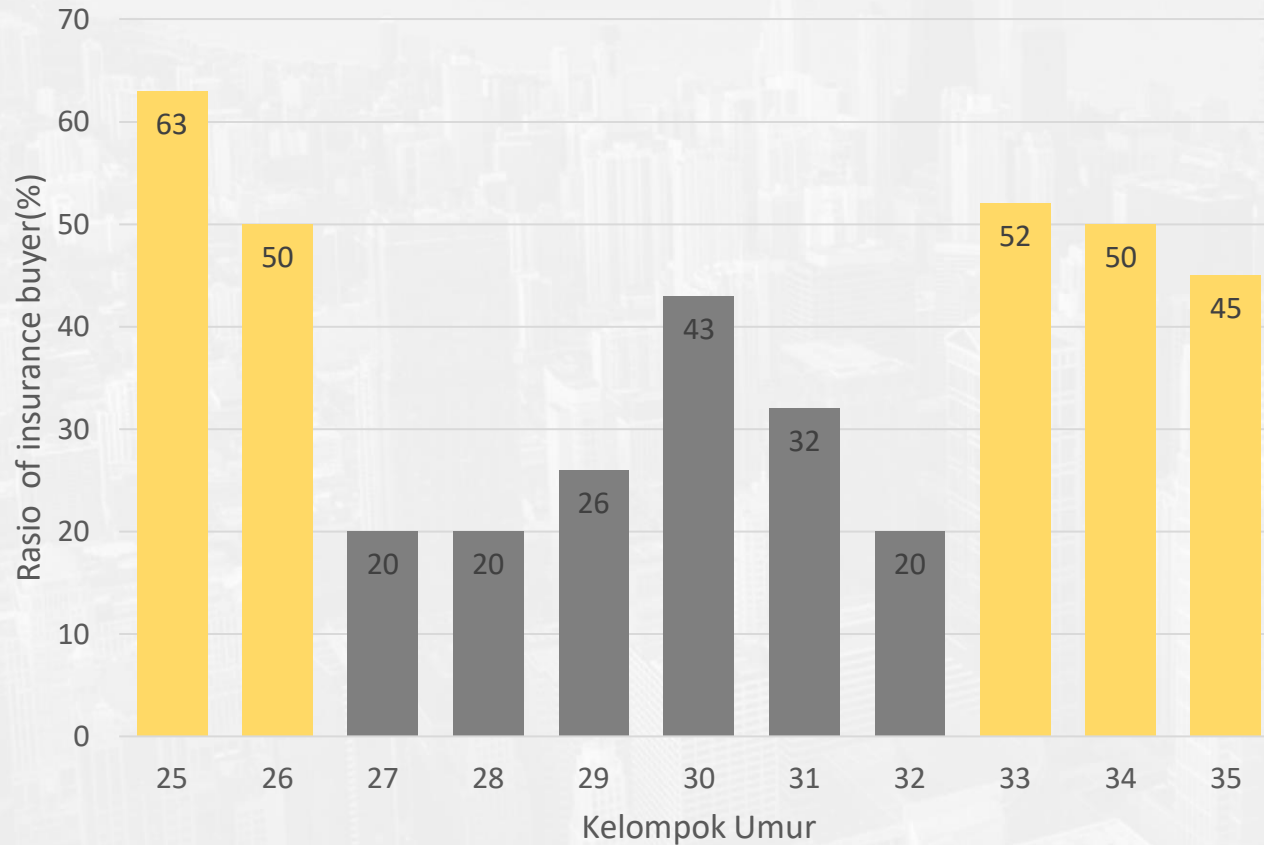


Travel Insurance cenderung dibeli oleh mereka dengan **Pendapatan Lebih Tinggi**

Kenaikan kecenderungan pembelian travel insurance terlihat pada mereka dengan pendapatan diatas 1.3M



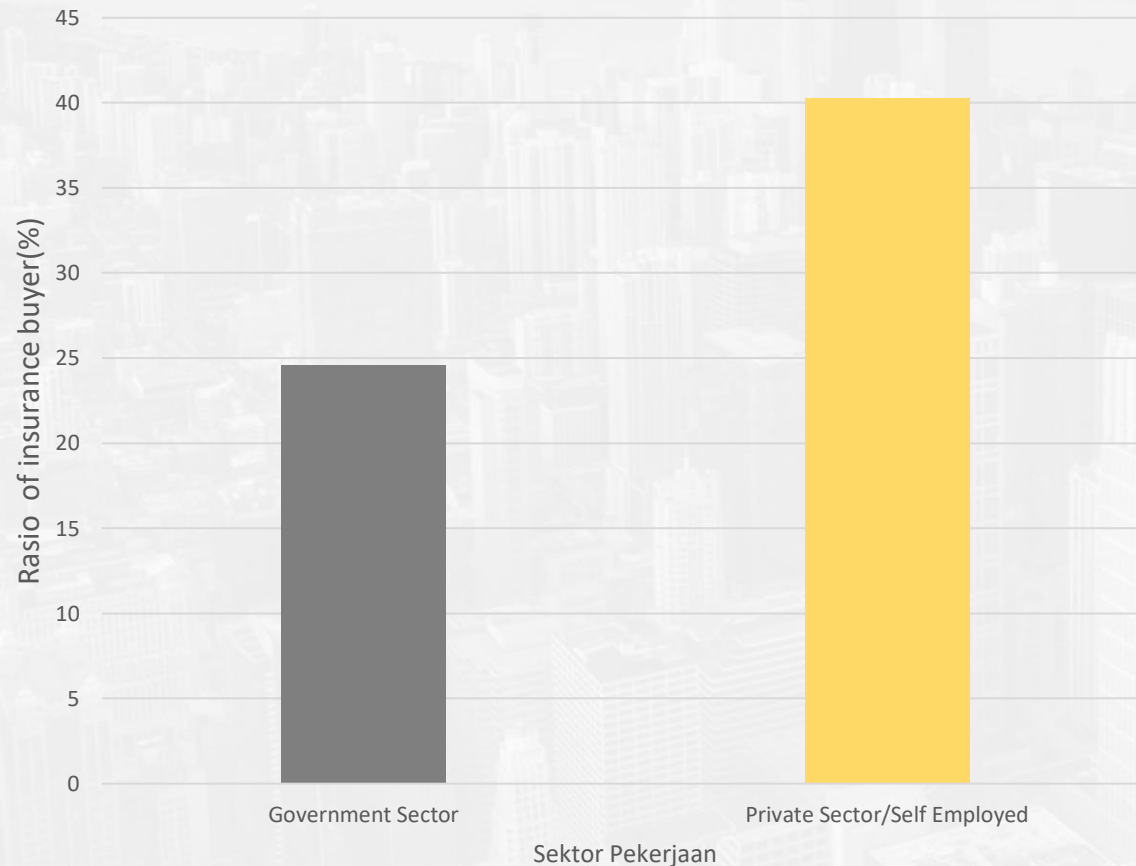
# Exploratory Data Analysis (EDA)



Travel Insurance cenderung dibeli oleh **demografi lebih muda atau lebih tua**

Rasio pembeli lebih tinggi ditemukan di **age group lebih muda ( $\leq 26$ )** dan **age group lebih tua ( $\geq 33$ )**

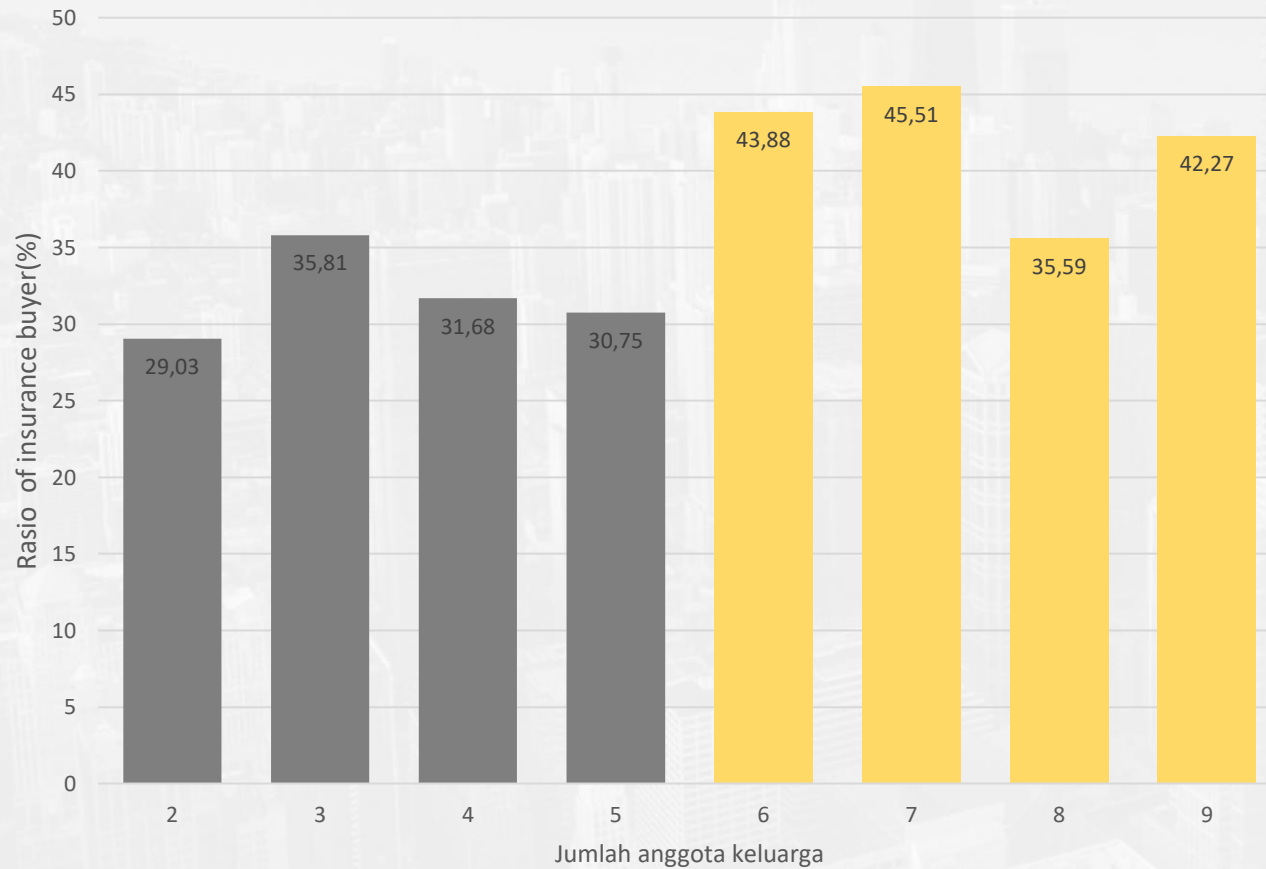
# Exploratory Data Analysis (EDA)



Travel Insurance cenderung dibeli oleh **private sector/self-employed**

Pekerja di **private sector/self employed** 1,6 kali lebih mungkin membeli travel insurance dibandingkan mereka yang bekerja di **sektor pemerintahan**

# Exploratory Data Analysis (EDA)



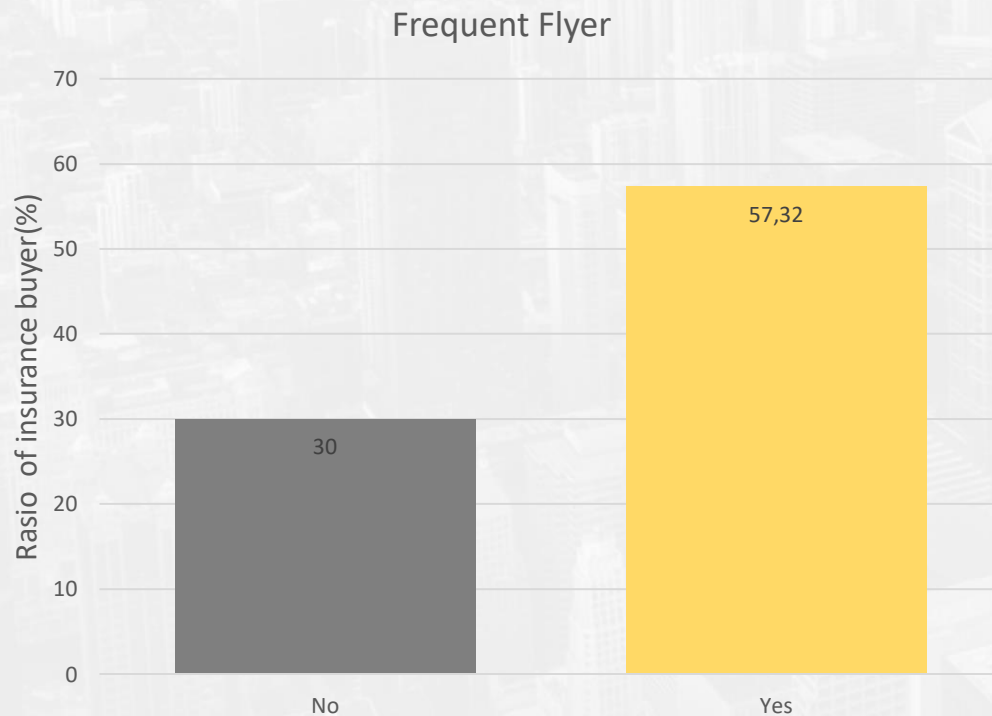
Travel Insurance cenderung dibeli travelers dengan **Jumlah Keluarga lebih besar**

Customer dengan jumlah anggota keluarga  $\geq 6$  kecenderungan pembelian lebih tinggi

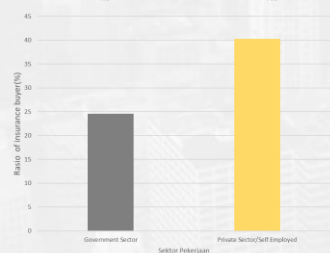
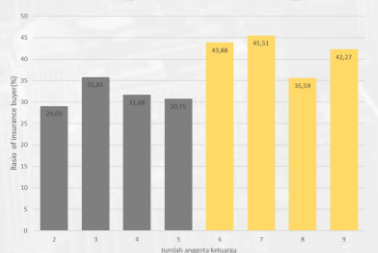
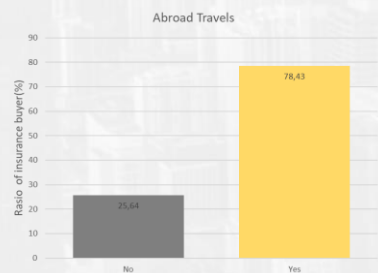
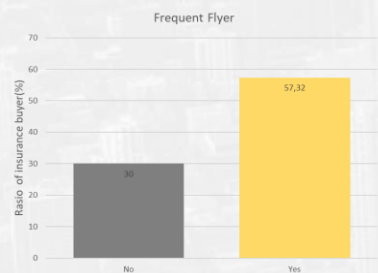
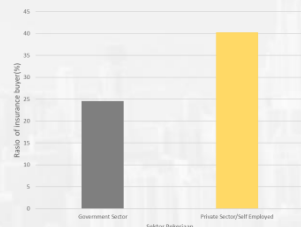
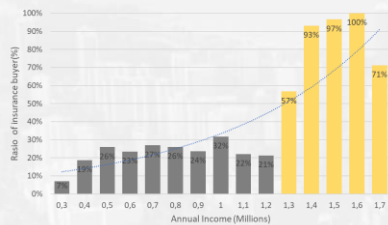
# Exploratory Data Analysis (EDA)

Asuransi cenderung dibeli oleh mereka yang lebih **“well-traveled”**

Mereka yang melakukan perjalanan **lebih sering dan lebih jauh (pernah pergi ke luar negeri)** memiliki **kecenderungan lebih tinggi dalam pembelian**







Dari beberapa hasil EDA diatas, kami melihat bahwa diperlukan **segmentasi pasar** pembeli Travel Insurance, yang selanjutnya dibuatkan Machine Learning Model.

Pemberlakuan segmentasi pasar diharapkan dapat mengurangi **marketing cost** serta **marketing dapat tertarget**

**Sebelum dilakukan modeling, diperlukan Data Pre-Processing, agar Decision-Making dan Business Recommendation tidak salah.**

# *Data Pre-Processing*



# Data Pre-Processing

- ✓ **No Null Values in dataset**  
Data tidak terdapat null values



- ✓ **Check & Handle Duplicates**  
Data before handled: 1987  
Data after handled: **1249**



- ✓ **Encoding**  
Label Encoding :  
**EmploymentType,**  
**GraduateOrNot,**  
**FrequentFlyer,**  
**EverTravelledAbroad**



- ✓ **Scaling & Transformation**  
Normalization & Standardization



# *Model & Evaluation*





# Model Comparison (Default)

| Evaluation           | Model               |               |               |             |                |             |             |
|----------------------|---------------------|---------------|---------------|-------------|----------------|-------------|-------------|
|                      | Logistic Regression | Decision Tree | Random Forest | SVM         | Neural Network | XGBOOST     | KNN         |
| Accuracy (Test Set)  | 0.70                | 0.77          | 0.66          | 0.78        | 0.72           | 0.82        | 0.70        |
| Precision (Test Set) | <b>0.70</b>         | <b>0.90</b>   | <b>0.56</b>   | <b>0.88</b> | <b>0.79</b>    | <b>0.92</b> | <b>0.65</b> |
| Recall (Test Set)    | 0.39                | 0.46          | 0.51          | 0.49        | 0.37           | 0.58        | 0.49        |
| F1-Score (Test Set)  | 0.50                | 0.61          | 0.53          | 0.62        | 0.51           | 0.71        | 0.56        |

**Model Evaluation** yang digunakan:

- Precision

**Precision** digunakan karena **tidak memperbolehkan nilai False Positive yang besar (Customer yang dianggap tidak berpotensi untuk membeli, namun dianggap berpotensi)**

Dari hasil perbandingan diatas, kami menentukan bahwa **Decision Tree** dan **XGBOOST** adalah model yang salah satunya akan digunakan, karena nilai **Precision yang tinggi**. Untuk menentukan nya, diperlukan **Hyperparameter Tuning** untuk menentukan apakah model ini sudah fit atau belum.

# Model after Hyperparameter Tuning

| Model Evaluation :<br>Precision | Decision Tree | XGBOOST |
|---------------------------------|---------------|---------|
| Train                           | <b>0,88</b>   | 0,87    |
| Test                            | <b>0,88</b>   | 0,91    |

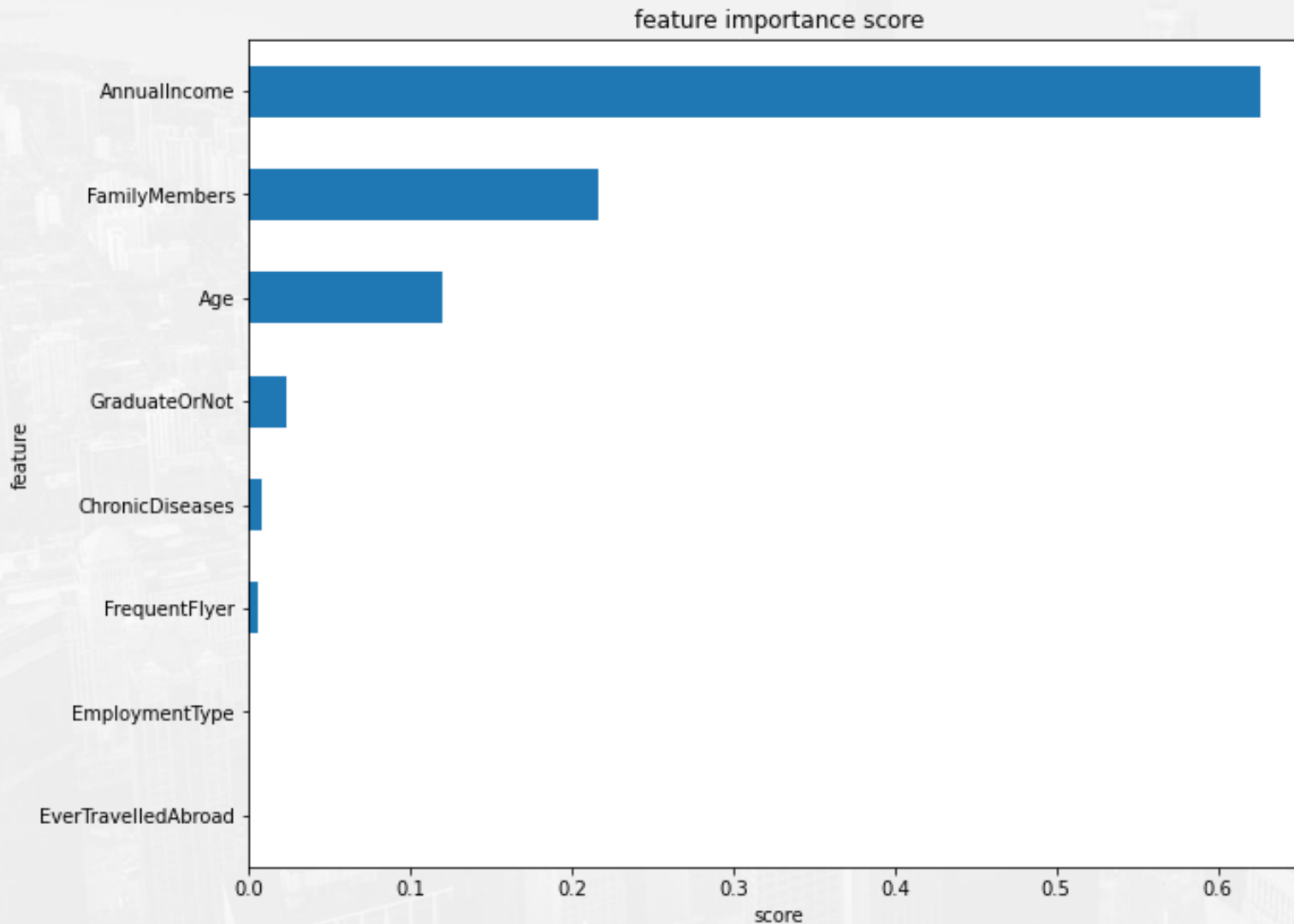
**Decision Tree** menjadi model yang dipilih karena :

- Model Decision Tree adalah yang paling **best fit**, karena nilai **Train dan Test dengan delta terkecil**
- Model sudah sesuai dengan target (>75%)

Beberapa trial yang dilakukan pada model Decision Tree yang best fit :

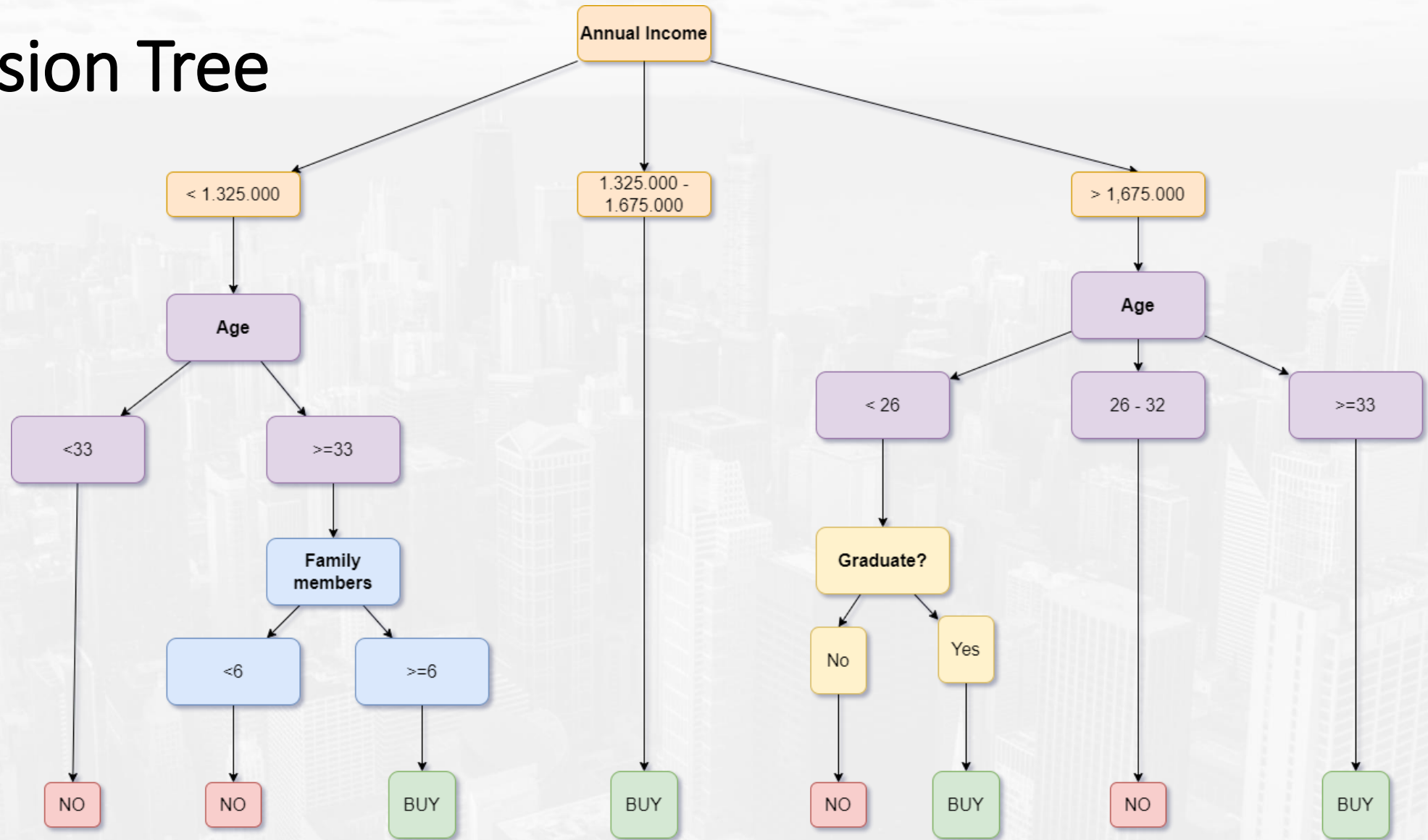
- Class Imbalance
- Hyperparameter tuning (max\_depth, criterion = 'entropy', min\_samples\_split, min\_samples\_leaf)

# Feature Importance



Dari hasil Feature Importance model Decision Tree, Fitur dengan kepentingan tertinggi adalah **Annual Income, Family Members, dan Age**

# Decision Tree





# *Business Insight & Recommendation*



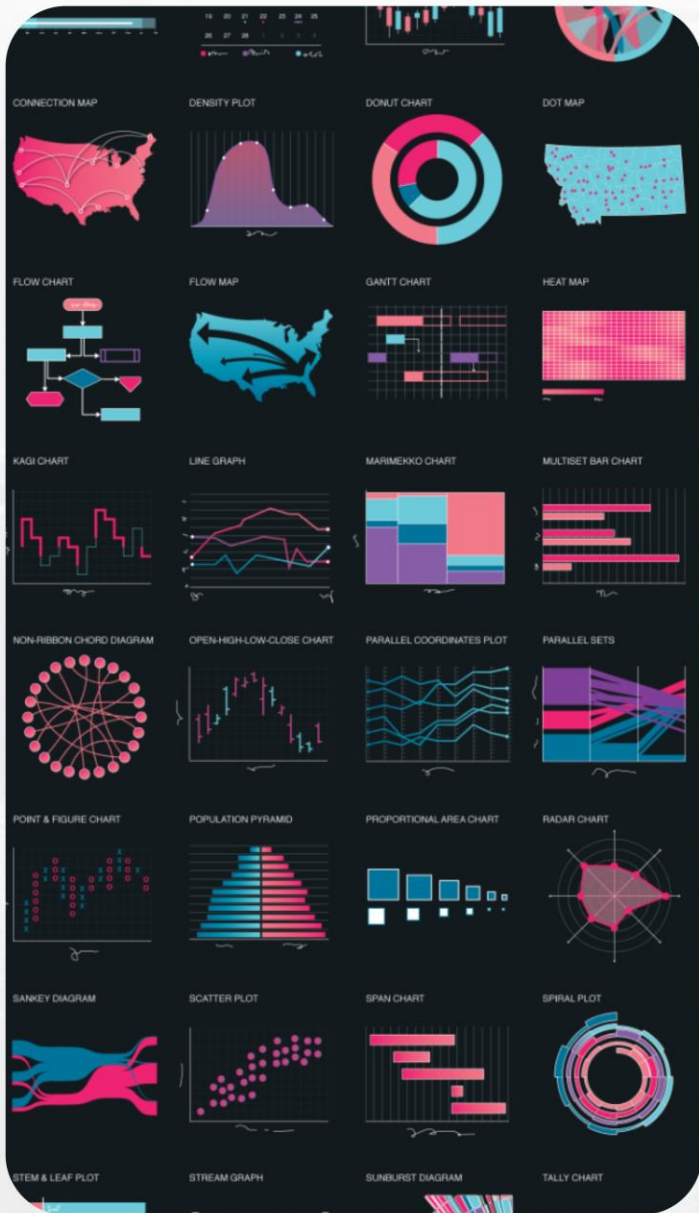
# Before & After Modelling



| Before using Model  | After using Model (Precision: 0,88)  |
|---|--|
| Customer contacted → 1987<br>Total insurance bought → 710 (35%)                   | Customer contacted → 806 (clean data)<br>Total insurance bought → <b>710 (88%)</b> |
| Marketing cost:<br>Rp 100.000,- x 1987 (potential customer)<br>= Rp 198.700.000,- | Marketing cost:<br>Rp 100.000,- x 806 (potential customer)<br>= Rp 80.600.000,-    |
| CAC=<br>Rp 198.700.000,- / 710<br>= Rp 279.859,-                                  | CAC=<br>Rp 80.600.000,- / 710<br>= Rp 113.521,-                                    |
| <b>60% decrease in CAC</b>  |  |

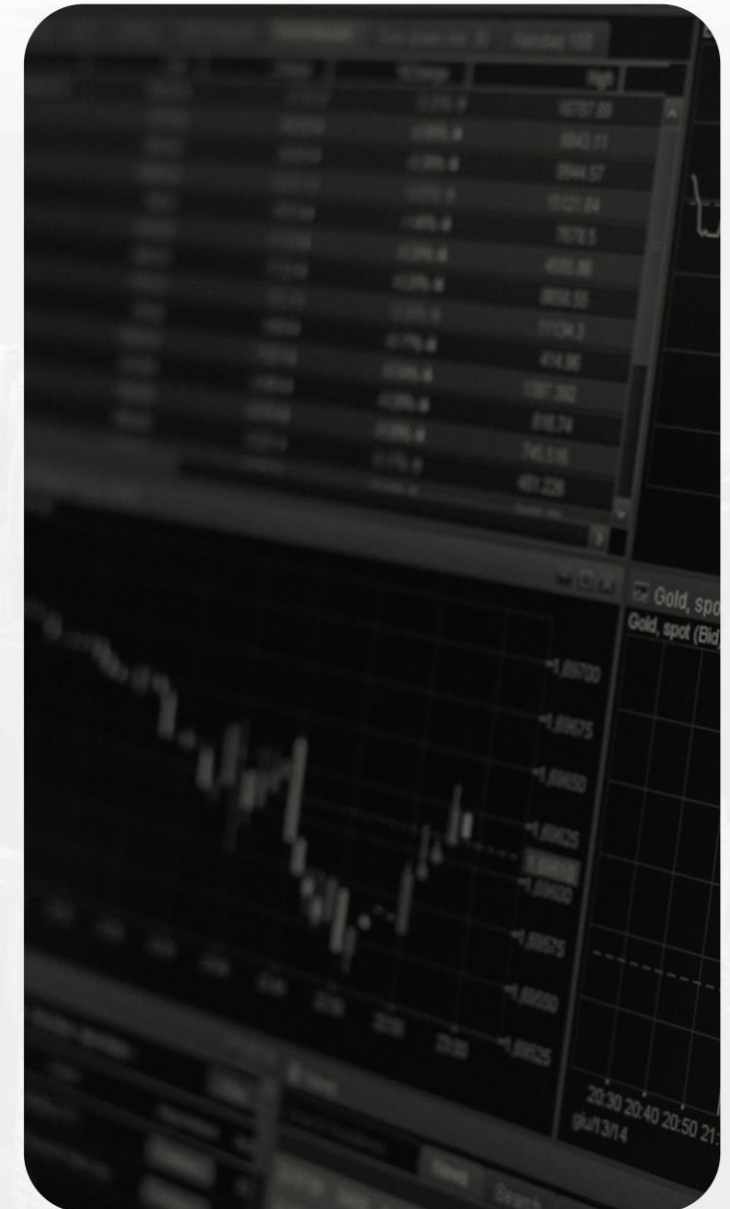
# Recommendation for Marketing Team

1. Melakukan **targeted marketing** berdasarkan hasil analisis:
  - A. Pendapatan tinggi (> 1.3Mil)
  - B. Anggota keluarga besar (6 orang/lebih)
  - C. Kelompok umur muda dan tua (26 kebawah dan 33 keatas)
  - D. Pekerja swasta/self-employed
  - E. Memiliki pengalaman bepergian lebih tinggi
2. **Membuat konten marketing yang dapat meningkatkan pengetahuan customer terhadap pentingnya travel insurance**
3. Menurunkan biaya marketing dengan mencari cara-cara alternatif lainnya, seperti **melakukan iklan di social media** dan **meningkatkan keragaman platform advertisement**



# Recommendation for Product Team

1. Membuat **alternatif** paket travel insurance untuk menjangkau segmentasi pasar seperti:
  - A. Membuat produk Travel Insurance dengan harga yang lebih rendah
  - B. Membuat produk asuransi add-on untuk government workers
  - C. Membuat alternative protection sesuai umur
  - D. Membuat program bundling/subscription bagi mereka yang sering bepergian
2. Menambah **benefit** asuransi untuk meningkatkan nilai produk, seperti:
  - A. Add-on family protection: tanggungan terhadap pasangan atau anak
  - B. Penambahan proteksi sesuai durasi perjalanan
  - C. Proteksi pembatalan, penundaan, keterlambatan perjalanan





# Thank You

hemapython<sup>®</sup>

