

# Hakuna Matata Travel Insurance

---

## Dokumen Laporan Final Project

(dipresentasikan setiap sesi mentoring)



# Latar Belakang Masalah (1-2 slide)

- Sebuah perusahaan travel insurance menemukan fakta bahwa hanya 35.73% customer yang membeli paket asuransi perjalanan mereka.
- Bagaimana cara untuk membantu team business dalam memprediksi customer yang akan tertarik membeli paket asuransi yang akan ditawarkan?
- Menggunakan analytical approach: membentuk sebuah model rekomendasi untuk mengklasifikasi para pelanggan berdasarkan karakteristik dan perjalanan mereka. Tujuannya agar terjadi peningkatan akurasi dalam memprediksi yang tertarik membeli paket asuransi.



# Pre-processing

## Satistik Deskriptif

A. Tipe data dan nama kolom:

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 1987 entries, 0 to 1986
```

```
Data columns (total 10 columns):
```

#	Column	Non-Null Count	Dtype
0	Unnamed: 0	1987 non-null	int64
1	Age	1987 non-null	int64
2	Employment Type	1987 non-null	object
3	GraduateOrNot	1987 non-null	object
4	AnnualIncome	1987 non-null	int64
5	FamilyMembers	1987 non-null	int64
6	ChronicDiseases	1987 non-null	int64
7	FrequentFlyer	1987 non-null	object
8	EverTravelledAbroad	1987 non-null	object
9	TravelInsurance	1987 non-null	int64

```
dtypes: int64(6), object(4)
```

```
memory usage: 155.4+ KB
```

```
None
```

# Cek nilai yang hilang

```
df.isnull().sum()
```

```
Unnamed: 0      0
Age              0
Employment Type  0
GraduateOrNot    0
AnnualIncome     0
FamilyMembers    0
ChronicDiseases  0
FrequentFlyer    0
EverTravelledAbroad 0
TravelInsurance  0
dtype: int64
```

# Cek duplikasi

```
df.duplicated().sum()
```

```
0
```

# Cek total unique value

```
dict = {}
for i in df.columns:
    print(f'{i} = {df[i].nunique()}')
```

```
Unnamed: 0 = 1987
```

```
Age = 11
```

```
Employment Type = 2
```

```
GraduateOrNot = 2
```

```
AnnualIncome = 30
```

```
FamilyMembers = 8
```

```
ChronicDiseases = 2
```

```
FrequentFlyer = 2
```

```
EverTravelledAbroad = 2
```

```
TravelInsurance = 2
```

# Pre-processing

## Statistik Deskriptif

A. Apakah ada kolom dengan tipe data kurang sesuai, atau nama kolom dan isinya kurang sesuai?

- \* Kolom 'GraduateOrNot', 'ChronicDiseases', 'FrequentFlyer', 'EverTravelledAbroad', dan 'TravelInsurance' adalah tipe data Boolean, jadi untuk keperluan EDA sebaiknya disamakan saja tipenya menjadi object atau integer agar mudah dikategorikan menjadi data numerikal dan kategorikal. Ini akan memudahkan pemahaman bahwa data numerikal adalah yang berisi angka (integer), dan data kategorikal berisi karakter/string (object).
- \* Kolom 'TravelInsurance' tidak diubah karena merupakan target dan nantinya dibutuhkan data numerikal saat modelling machine learning.

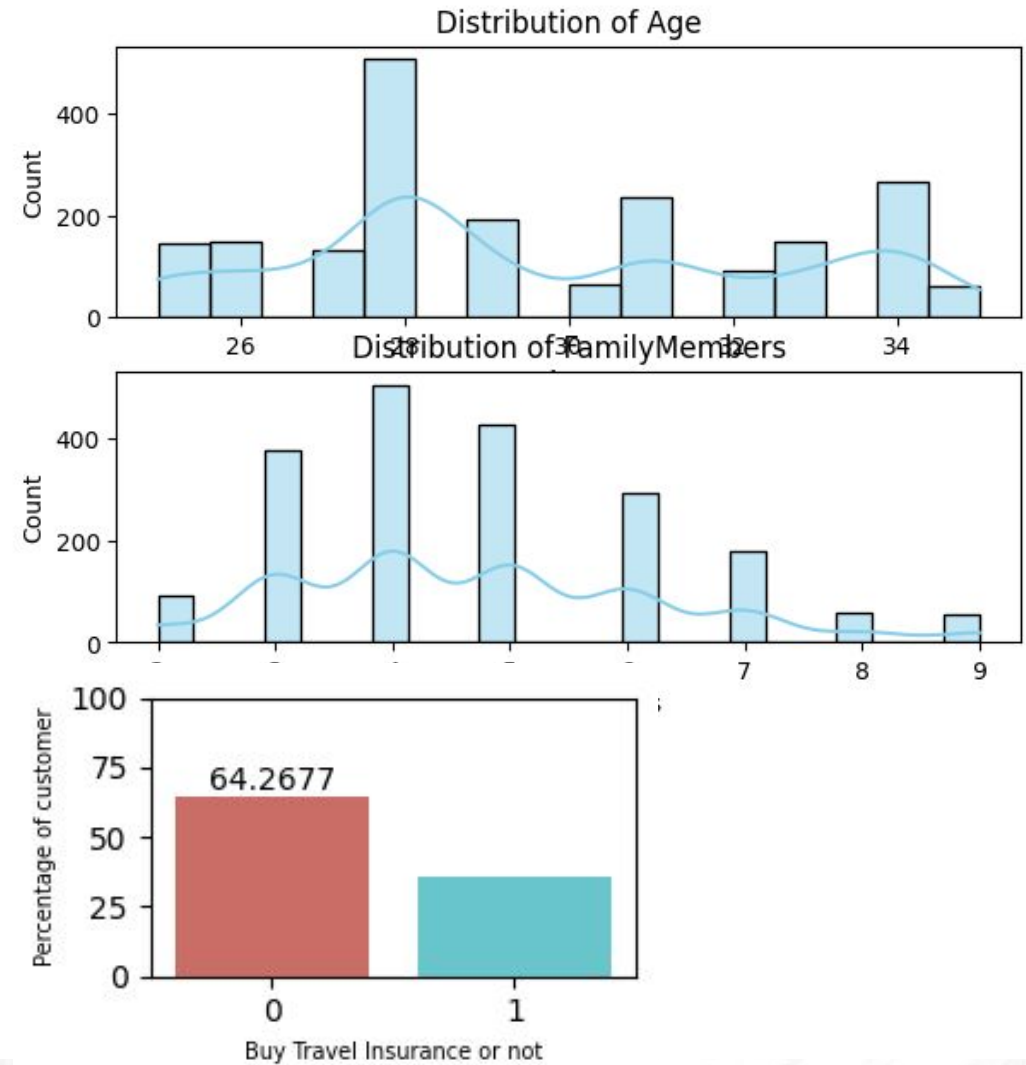
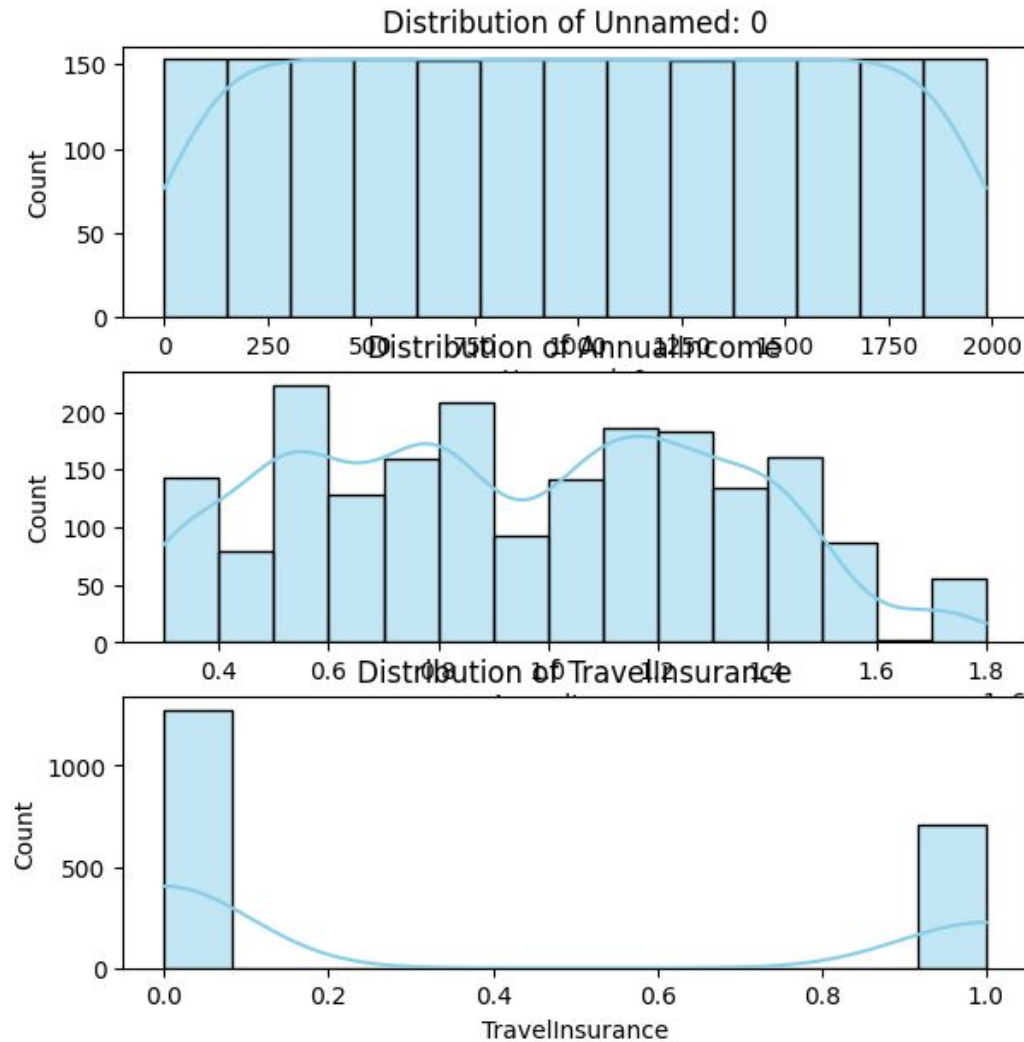
B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?  
Tidak ada kolom yang memiliki nilai kosong dalam dataset ini.

C. Apakah ada kolom yang memiliki nilai summary agak aneh?  
(min/mean/median/max/unique/top/freq)

- \* Rentang umur pelanggan adalah 25-35, termasuk golongan usia produktif yang muda.
- \* Pelanggan didominasi oleh karyawan swasta dan wiraswasta, lulusan universitas, tidak mengidap penyakit kronis, tidak sering bepergian dengan pesawat dan belum pernah traveling keluar negeri.
- \* Pelanggan dengan frekuensi terbanyak berusia 28 tahun, dengan Pendapatan Tahunan 800.000, dan memiliki 4 Anggota Keluarga
- \* Pelanggan yang membeli asuransi perjalanan hanya 35,73%

# Pre-processing

## Analisis Univariat

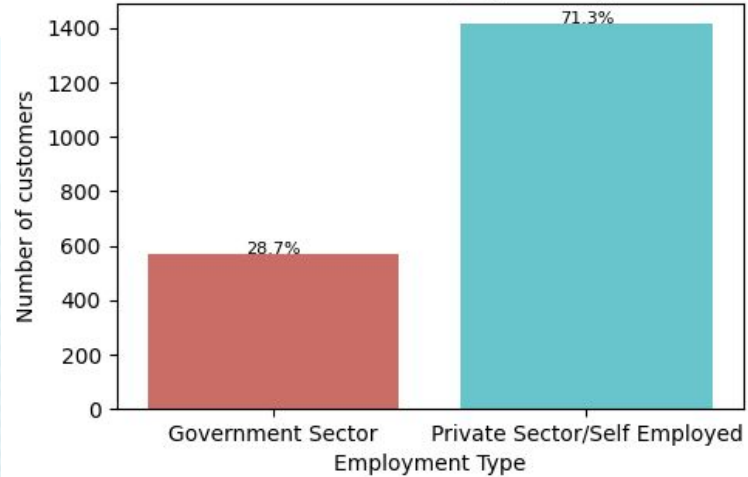




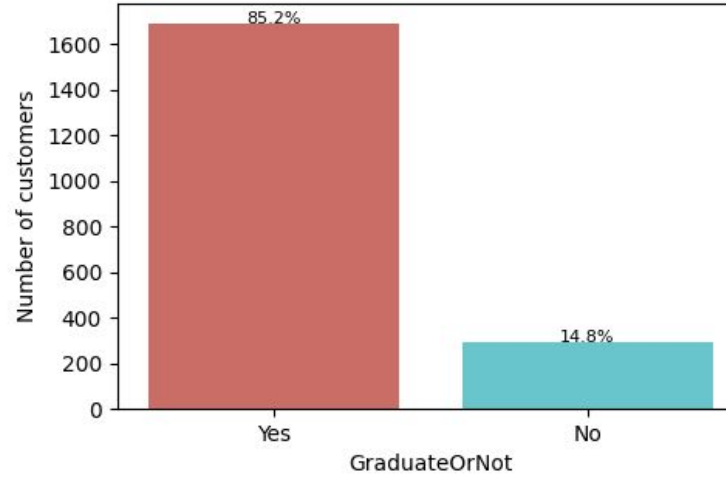
# Pre-processing

## Analisis Univariat

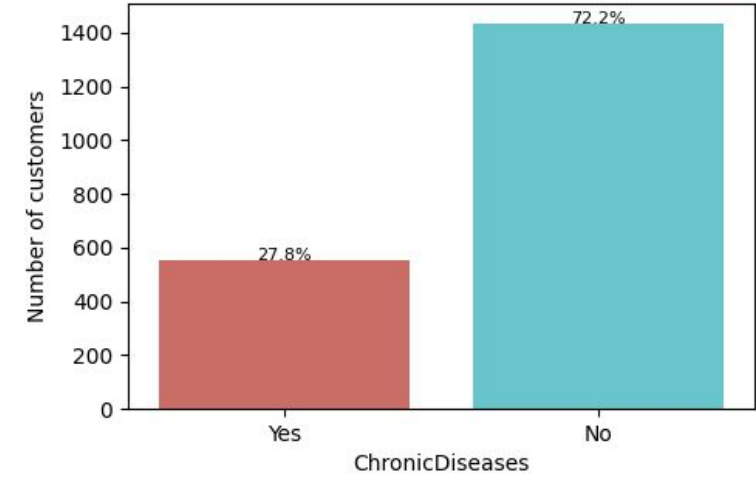
Distribution of Employment Type



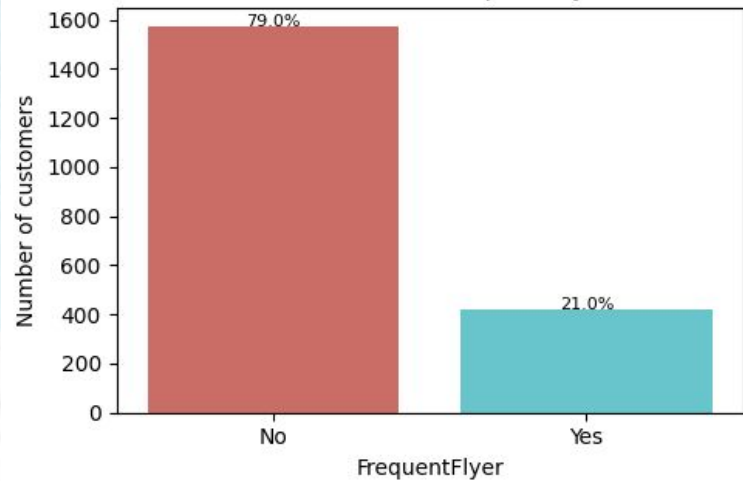
Distribution of GraduateOrNot



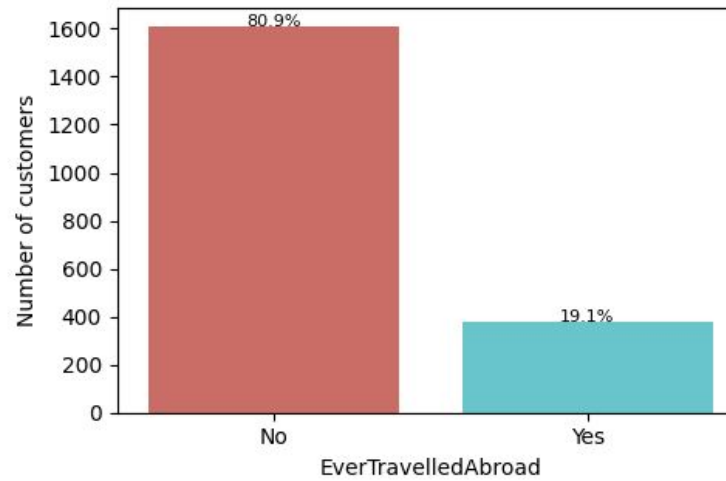
Distribution of ChronicDiseases



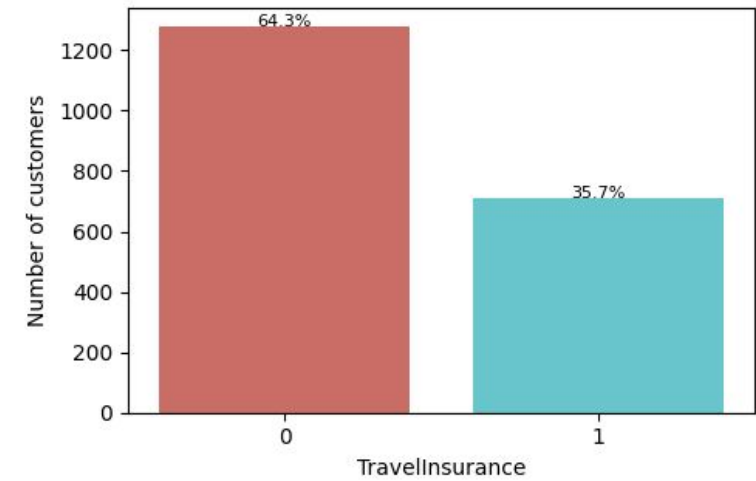
Distribution of FrequentFlyer



Distribution of EverTravelledAbroad



Distribution of TravellInsurance



# Pre-processing

## Analisis Univariat

### A. Fitur Numerikal

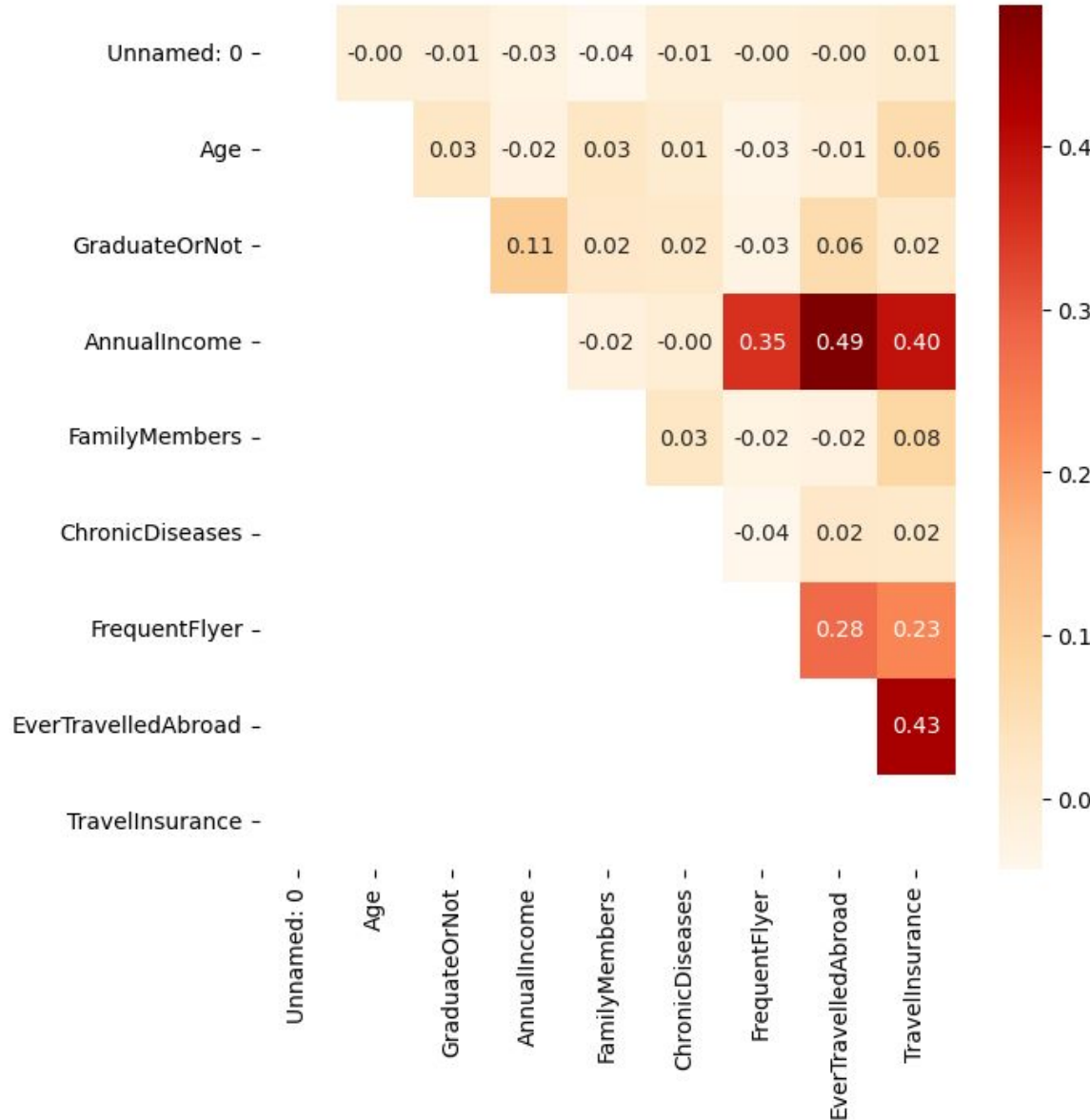
- \* Distribusi umur bervariasi, dengan mean, median, dan modus yang mirip (29.650226, 29, dan 28)
- \* Distribusi annual income bervariasi namun merata, dengan mean = 932,763 dan median = 900,000
- \* Distribusi anggota keluarga cenderung positif (positively skewed), dengan rata-rata (median) 5 orang. Anggota keluarga yang paling banyak muncul adalah 4 orang. Tiap pelanggan memiliki setidaknya 2 orang anggota keluarga
- \* Lebih dari 60% pelanggan tidak membeli travel insurance
- \* Tidak ada outlier

### B. Fitur Kategorikal

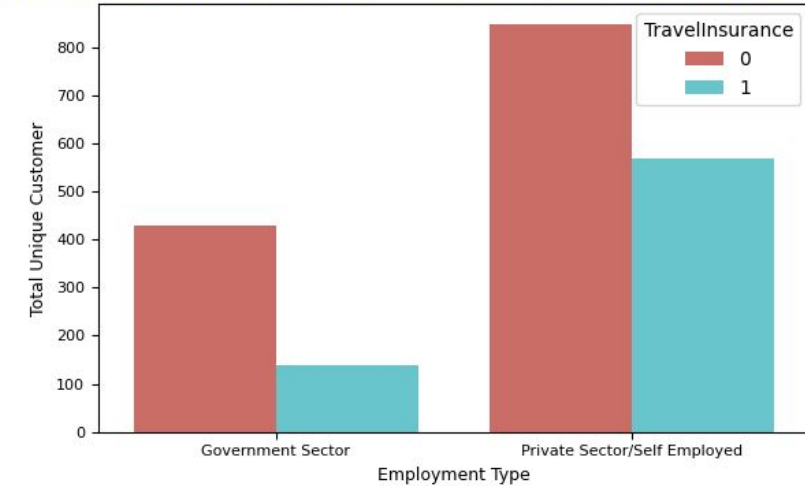
- \* Lebih dari 70% pelanggan adalah karyawan swasta dan wiraswasta
- \* Sekitar 85% pelanggan adalah lulusan perguruan tinggi
- \* 27.8% pelanggan mengidap penyakit kronis
- \* Hanya 21% pelanggan yang sering bepergian dengan pesawat
- \* Hanya 19% pelanggan yang pernah traveling keluar negeri

# Pre-processing

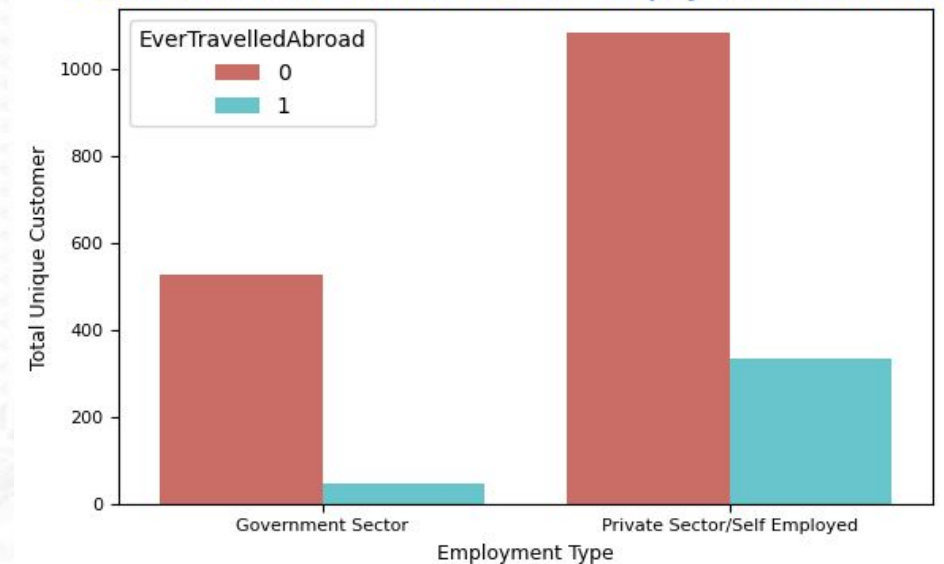
## Analisis Multivariat



Majority of customers who bought travel insurance work in Private Sector/Self Employed



Customers who work in Private Sector/Self Employed often travel abroad

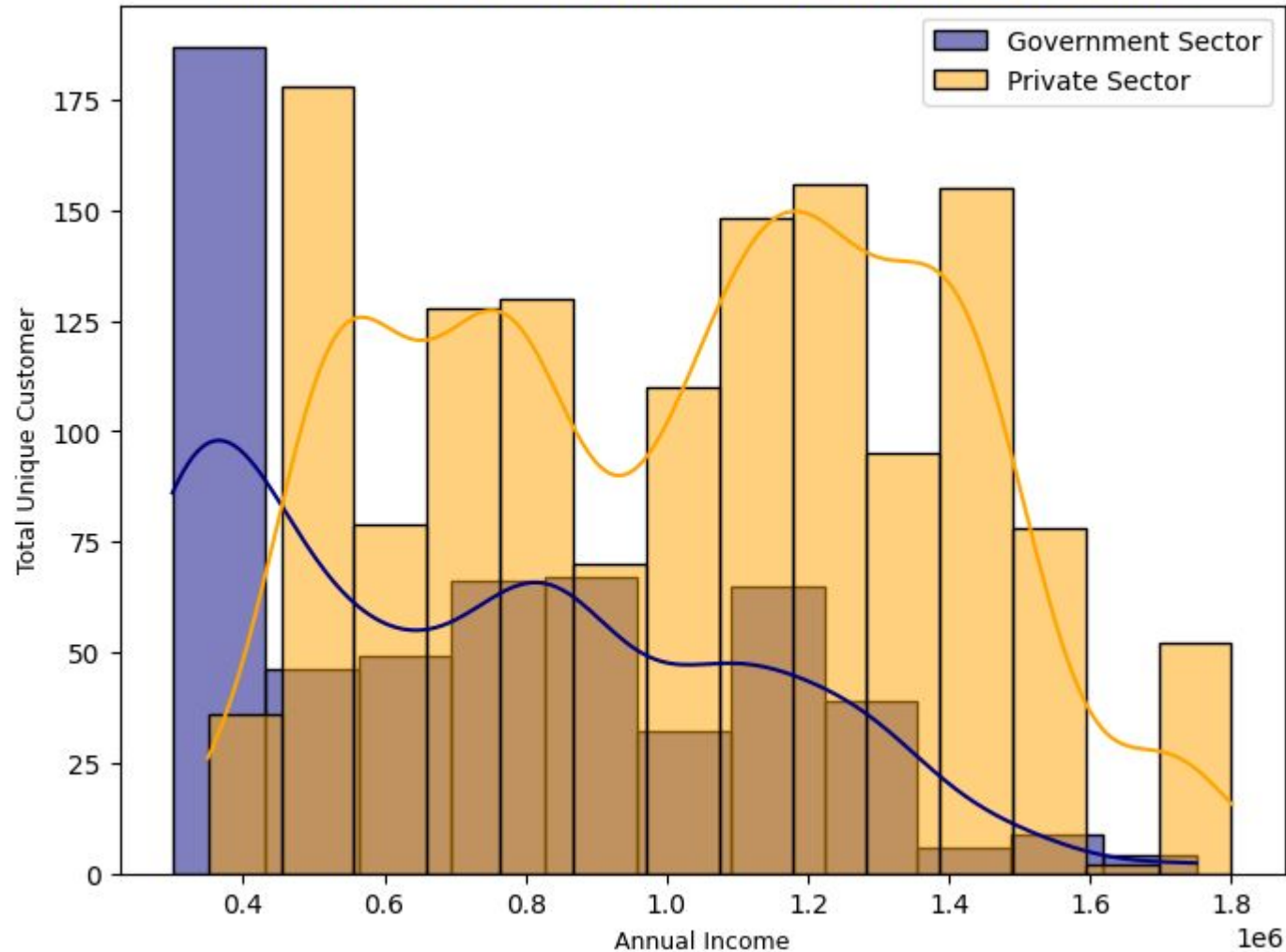




# Pre-processing

## Analisis Multivariat

There are more Private Sector workers who have higher income



	Employment Type	Average Annual Income
0	Government Sector	700000.0
1	Private Sector/Self Employed	1050000.0

# Pre-processing

## Analisis Multivariat

### A. Korelasi antara Feature Numerik dan Target ('TravelInsurance'):

- \* Target memiliki korelasi positif paling tinggi dengan EverTravelledAbroad (0.43). Pelanggan yang pernah keluar negeri akan cenderung membeli travel insurance.
- \* Korelasi antara target dengan Annual Income cukup tinggi (0.4). Semakin tinggi pendapatan pelanggan, semakin besar kemungkinan pelanggan membeli travel insurance.
- \* Potensi pelanggan yang FrequentFlyer untuk membeli Travel Insurance cukup potensial (0,23). Semakin sering berpergian pelanggan akan lebih memiliki kesadaran pentingnya memiliki sebuah Travel Insurance.

### B. Korelasi antar Feature:

- \* Korelasi antara Annual Income dengan EverTravelledAbroad cukup tinggi (0.49), menunjukkan bahwa pelanggan dengan pendapatan yang tinggi pernah berpergian keluar negeri.
- \* Korelasi antara Annual Income dengan FrequentFlyer cukup tinggi (0.35), menunjukkan bahwa pelanggan dengan pendapatan yang tinggi sering berpergian dengan pesawat.

# Pre-processing

## Analisis Multivariat

### C. Korelasi antara Feature Kategorik dan Target ('TravelInsurance'):

Pelanggan yang bekerja di bidang swasta dan wiraswasta cenderung membeli travel insurance.

### D. Korelasi antara Feature Kategorik dan Feature lainnya:

- \* Pelanggan yang pernah keluar negeri didominasi oleh karyawan swasta dan wiraswasta.
- \* Pelanggan yang bekerja di pemerintah kebanyakan memiliki pendapatan di bawah 400,000, semakin tinggi pendapatan semakin sedikit jumlah orangnya. Dapat dilihat juga rata-rata pendapatannya di bawah karyawan yang bekerja di sektor swasta. Sedangkan Pelanggan yang bekerja di sektor swasta memiliki rata-rata pendapatan lebih tinggi, dengan jumlah karyawan yang memiliki pendapatan di atas 450,000 lebih banyak.



- \* Fokuskan strategi marketing pada pelanggan yang pernah bepergian ke luar negeri.
- \* Lakukan segmentasi pelanggan berdasarkan kategori pendapatan tahunan. Misalnya, tentukan rentang pendapatan rendah, menengah, dan tinggi. Sesuaikan penawaran dan premi travel insurance sesuai dengan segmentasi ini. Pelanggan dengan pendapatan lebih tinggi mungkin lebih bersedia untuk membayar premi yang lebih tinggi untuk manfaat tambahan. Berikan diskon atau insentif khusus kepada pelanggan dengan pendapatan tinggi.
- \* Tawarkan paket asuransi perjalanan premium dengan manfaat tambahan kepada pelanggan dengan pendapatan tinggi yang memiliki riwayat perjalanan internasional ('EverTravelledAbroad' = 'Yes') dan sering bepergian dengan pesawat ('FrequentFlyer' = 'Yes').
- \* Fokuskan strategi marketing kepada pelanggan yang bekerja di sektor swasta dan wiraswasta. Perusahaan dapat melakukan hal ini dengan memanfaatkan media-media yang diminati oleh segmen pasar tersebut, seperti media sosial atau komunitas online.
- \* Perusahaan dapat menawarkan paket asuransi perjalanan terbaru dengan harga yang lebih kompetitif untuk menarik minat pelanggan yang bekerja di sektor pemerintah.

# Pembagian Tugas

Seluruh stage ini dikerjakan bersama-sama secara keseluruhan dengan beberapa poin lain dilakukan pembagian tugas, yakni:

Ahmad Faqih Ulumuddin	: Memfasilitasi teman-teman kelompok di discord grup saat sesi mentoring.
Ana Azzahra	: Membuat Notulensi.
Athiya Fathinati Anindya	: Share Screen dan presentasi ke mentor saat mentoring.
Esa Risa Rouli Sekar Andonowari	: Membuat git.
Jerio Benediktus Rumagit	: Membuat laporan.
Mutiara Citra Sari	: Membuat laporan.
Nicken Shidqia Nurahman	: Membuat notulensi.