

Group 1

Dokumen Laporan Final Project



Dataset yang Dipakai

Travel Insurance Prediction

<https://www.kaggle.com/datasets/tejashvi14/travel-insurance-prediction-data>

Variabel	Definisi
Age	Umur customer
Employment Type	Sektor dimana customer bekerja
<u>GraduateOrNot</u>	Apakah pelanggan adalah lulusan perguruan tinggi atau tidak
<u>AnnualIncome</u>	Pendapatan tahunan customer
<u>FamilyMembers</u>	Jumlah anggota keluarga yang tinggal bersama customer
<u>ChronicDiseases</u>	Apakah customer menderita penyakit kronis atau kondisi seperti diabetes / bp tinggi atau asthma, dll.
<u>FrequentFlyer</u>	Apakah customer sering memesan tiket pesawat
<u>EverTravelledAbroad</u>	Apakah customer pernah bepergian ke luar negeri.
<u>TravelInsurance</u>	Apakah customer membeli asuransi perjalanan atau tidak

1. Descriptive Statistics

Berikut adalah sampel data yang dipakai Travel Insurance

```
# Melihat sample data  
df.head()
```

	Unnamed: 0	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
0	0	31	Government Sector	Yes	400000	6	1	No	No	0
1	1	31	Private Sector/Self Employed	Yes	1250000	7	0	No	No	0
2	2	34	Private Sector/Self Employed	Yes	500000	4	1	No	No	1
3	3	28	Private Sector/Self Employed	Yes	700000	3	1	No	No	0
4	4	28	Private Sector/Self Employed	Yes	700000	8	1	Yes	No	0

Terdapat unnamed column yang tidak digunakan sehingga dilakukan fungsi drop pada kolom tersebut

```
[29] # Rename unnamed menjadi index  
df = df.rename(columns = {'Unnamed: 0' : 'index'})  
df.head()
```

	index	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
0	0	31	Government Sector	Yes	400000	6	1	No	No	0
1	1	31	Private Sector/Self Employed	Yes	1250000	7	0	No	No	0
2	2	34	Private Sector/Self Employed	Yes	500000	4	1	No	No	1
3	3	28	Private Sector/Self Employed	Yes	700000	3	1	No	No	0
4	4	28	Private Sector/Self Employed	Yes	700000	8	1	Yes	No	0

1. Descriptive Statistics

Untuk mengetahui beberapa informasi dari dataset yang ada, dijalankan fungsi info()

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

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

#	Column	Non-Null Count	Dtype
0	index	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
```

Informasi yang bisa diperoleh :

1. Data terdiri dari 9 kolom fitur, 1 kolom target, dan 1987 baris data.
2. Kolom age, annual income, family members, chronic diseases dan travel insurance bertipe integer.
3. Kolom Employment type, graduate or not, frequent flyer, dan ever travelled abroad bertipe object.
4. Keempat kolom yang bertipe object pada point nomor 3 merupakan data Boolean yang bisa dilakukan feature encoding pada stage selanjutnya.
5. Pada dataset ini, tidak ditemukan data null maupun NaN.

1. Descriptive Statistics

Cek data duplikat

```
# Cek data duplikat
df.duplicated(subset = ['Age', 'Employment Type', 'GraduateOrNot', 'AnnualIncome', 'FamilyMembers', 'ChronicDiseases', 'FrequentFlyer', 'EverTravelledAbroad', 'TravelInsurance']).sum()
```

738

Pada dataset ini kita mengecualikan pengecekan kolom index dikarenakan kolom index bukan merupakan bagian dari kolom yang mengidentifikasi customer maka kolom ini akan berpengaruh dalam pengecekan duplikat pada data dan hasilnya ditemukan sebanyak 738 baris data duplikat. Data duplikat tersebut tidak di drop terlebih dahulu pada step ini, dan akan dilakukan pada stage 2 pada saat tahap data preprocessing.

1. Descriptive Statistics

Melakukan fungsi describe() untuk mengetahui beberapa informasi data kolom numerik

df.describe()					
	Age	AnnualIncome	FamilyMembers	ChronicDiseases	TravelInsurance
count	1987.000000	1.987000e+03	1987.000000	1987.000000	1987.000000
mean	29.650226	9.327630e+05	4.752894	0.277806	0.357323
std	2.913308	3.768557e+05	1.609650	0.448030	0.479332
min	25.000000	3.000000e+05	2.000000	0.000000	0.000000
25%	28.000000	6.000000e+05	4.000000	0.000000	0.000000
50%	29.000000	9.000000e+05	5.000000	0.000000	0.000000
75%	32.000000	1.250000e+06	6.000000	1.000000	1.000000
max	35.000000	1.800000e+06	9.000000	1.000000	1.000000

Informasi yang diperoleh :

1. Range data min dan max masih bisa dikatakan wajar atau belum terlihat adanya data yang berbeda signifikan di setiap kolom
2. Nilai mean dan median tidak terlampau jauh untuk setiap kolomnya
3. Pada kolom age dan annual income nilai mean > median, sehingga datanya positively skewed.
4. Sedangkan data family member negatively skewed karena nilai mean < median.

1. Descriptive Statistics

Melakukan fungsi describe() khusus untuk data kolom kategorik

```
cat = ['Employment Type', 'GraduateOrNot', 'FrequentFlyer', 'EverTravelledAbroad']
df[cat].describe()
```

	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

Informasi yang diperoleh :

1. Mayoritas customer pada dataset ini berasal dari private sector / self employed dan graduate. Namun, kebanyakan customer bukan frequent flyer dan never travelled abroad. (lebih dari 70%).

1. Descriptive Statistics

Summary Descriptive Statistics

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

Answer : Terdapat kolom unnamed yang seharusnya direname menjadi index atau id.

B. Apakah ada kolom yang memiliki nilai kosong? Jika ada, apa saja?

Answer : Tidak terdapat kolom yang memiliki nilai kosong

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

Answer :

- Pada nilai max/unique/top/freq memiliki nilai yang wajar.
- Pada kolom age dan annual income nilai mean > median, sehingga datanya positively skewed.
- Sedangkan data family member negatively skewed karena nilai mean < median.

2. Univariate Analysis

Pada univariate analysis dataset dipisahkan berdasarkan tipe data nya terlebih dahulu

```
#Memisahkan data numerik
num = ['Age', 'AnnualIncome', 'FamilyMembers']
num

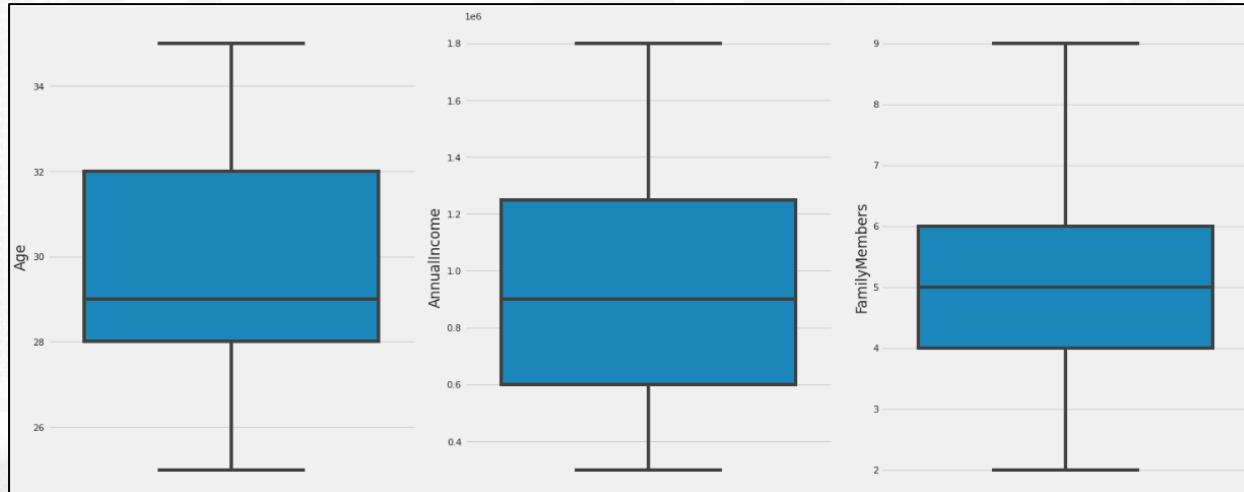
['Age', 'AnnualIncome', 'FamilyMembers']
```

```
#Memisahkan data kategorik
cat = ['Employment Type', 'GraduateOrNot', 'FrequentFlyer', 'ChronicDiseases', 'EverTravelledAbroad', 'TravelInsurance']
cat

['Employment Type',
 'GraduateOrNot',
 'FrequentFlyer',
 'ChronicDiseases',
 'EverTravelledAbroad',
 'TravelInsurance']
```

2. Univariate Analysis

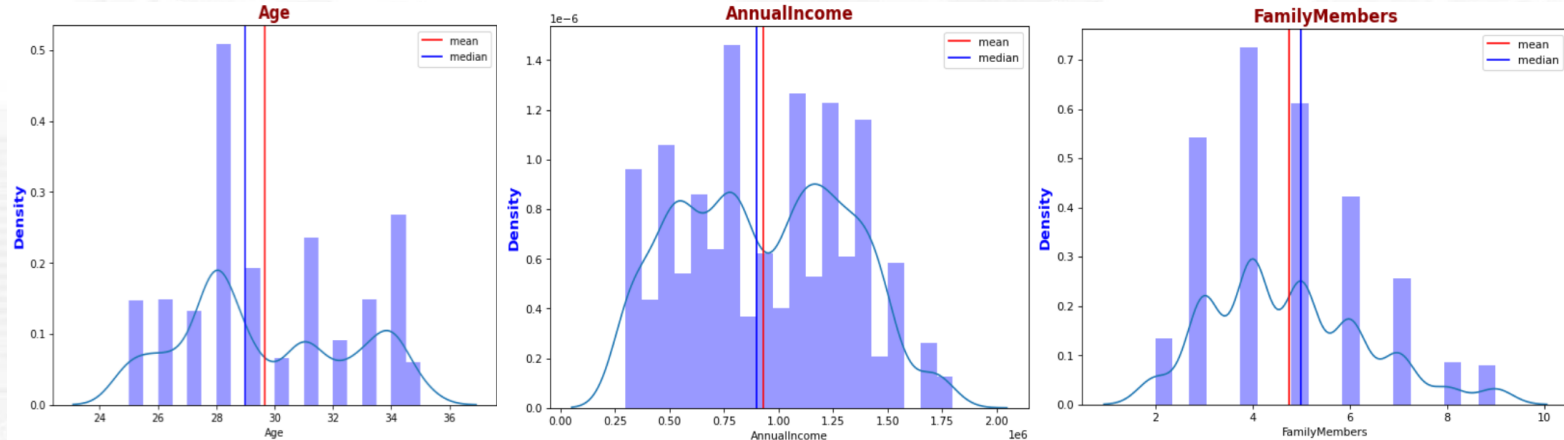
Univariate Analysis Data Numerik - Boxplot



Berdasarkan grafik boxplot diatas dapat diketahui bahwa data dari variabel yang berjenis numerikal '**Age**', '**AnnualIncome**', dan '**FamilyMember**' tidak memiliki outliers.

2. Univariate Analysis

Univariate Analysis Data Numerik – Distribution Plot



Informasi yang diperoleh:

- Variabel Age dan AnnualIncome bersifat positevely skew karena $\text{mean} > \text{median}$ dan plot menjorok kekanan sedangkan FamilyMember bersifat negatively skew karena $\text{median} > \text{mean}$ dan plot menjorok ke kekiri.
- Variabel Age dan AnnualIncome persebaran datanya simetris karena skewness value berada diantara -0,5 dan 0,5 sedangkan variabel FamilyMembers persebaran datanya moderately skew karena skewness value berada diantara 0.5 dan 1.
- Variabel Annual income merupakan multimodal karena memiliki lebih dari 1 mode.

Skew Score

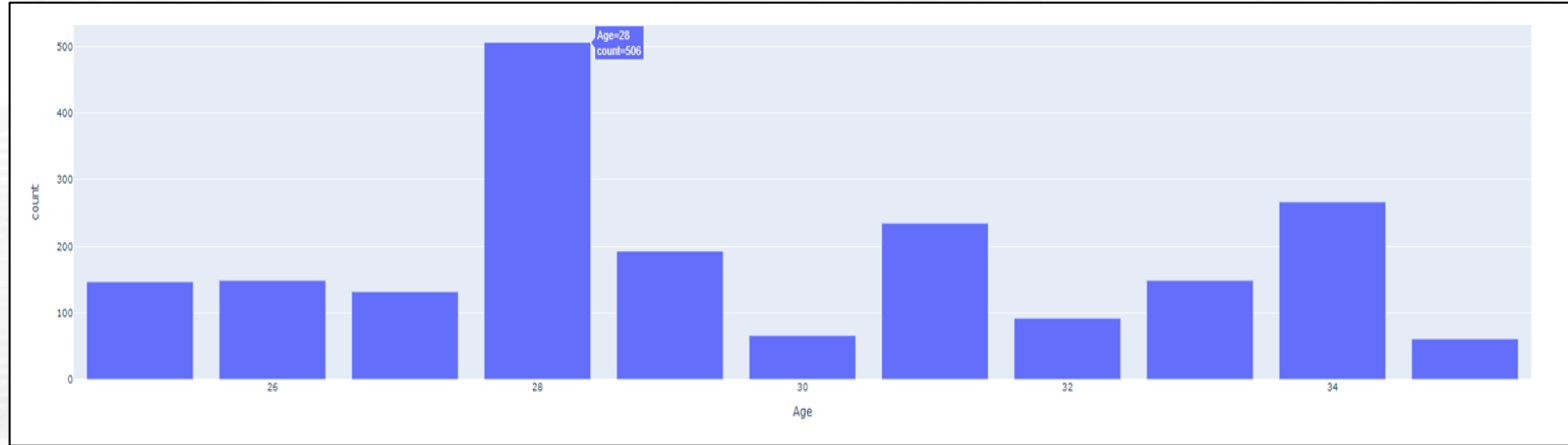
Skewness Age : 0.23962475139295264

Skewness AnnualIncome : 0.07841699043593056

Skewness FamilyMembers : 0.5611969123054563

2. Univariate Analysis

Univariate Analysis Data Numerik – Histogram (1/3)



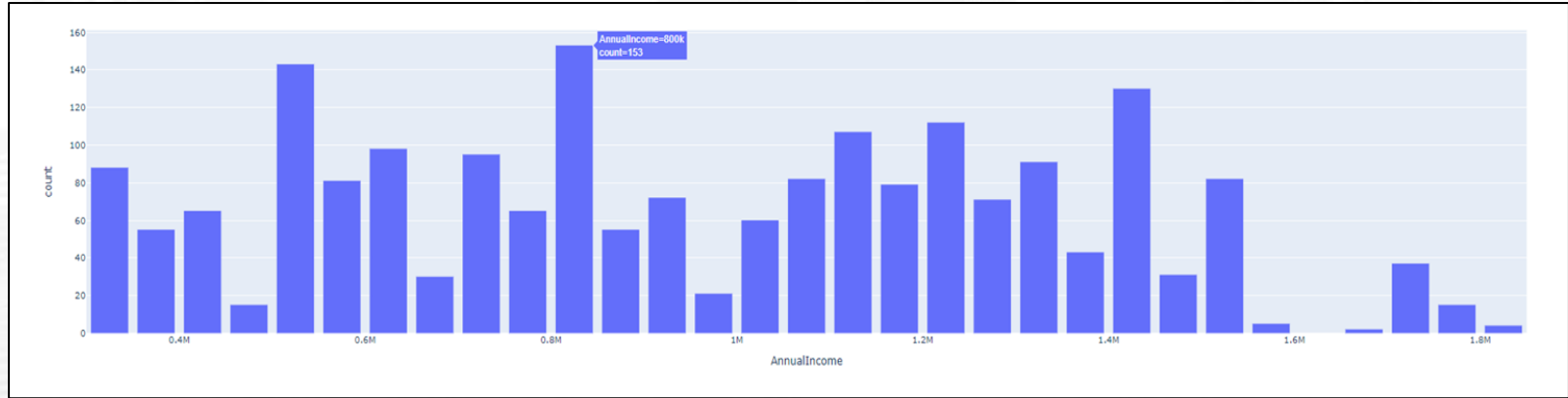
	Age	count	percent
0	28	506	25.465526
1	34	266	13.387016
2	31	234	11.776548
3	29	192	9.662808
4	33	148	7.448415

Informasi yang diperoleh:

1. Pada histogram tersebut mayoritas customer berusia 28 tahun dengan presentasi 25,4% dari keseluruhan data

2. Univariate Analysis

Univariate Analysis Data Numerik – Histogram (2/3)



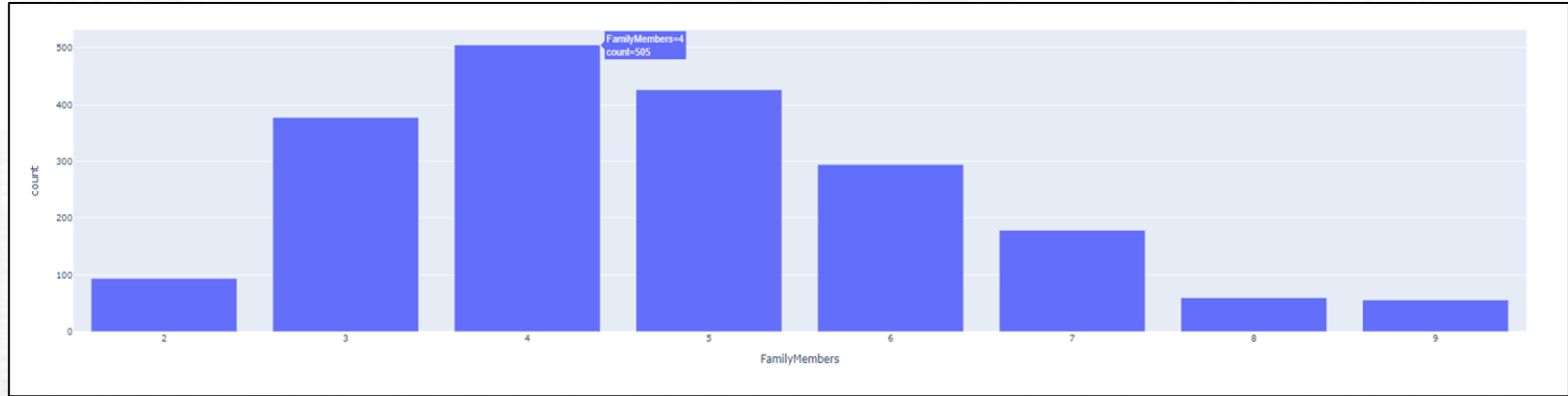
	AnnualIncome	count	percent
0	800000	153	7.700050
1	500000	143	7.196779
2	1400000	130	6.542526
3	1200000	112	5.636638
4	1100000	107	5.385003

Informasi yang diperoleh:

1. Tiga annual income tertinggi customer pada dataset ini adalah 0.5, 0.8, dan 1.4 INR
2. Pada histogram tersebut mayoritas customer memiliki annual income 800.000 dengan presentasi 7,7% dari keseluruhan data

2. Univariate Analysis

Univariate Analysis Data Numerik – Histogram (3/3)



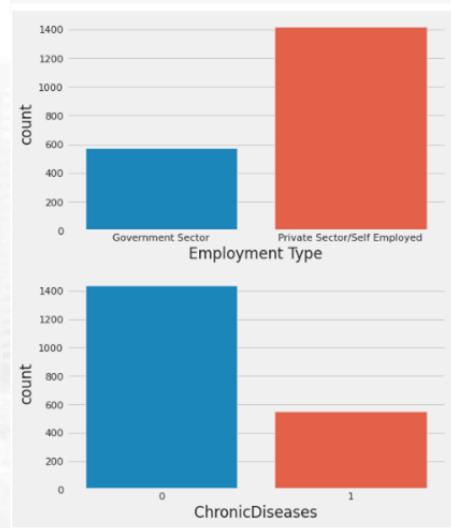
	FamilyMembers	count	percent
0	4	505	25.415199
1	5	426	21.439356
2	3	377	18.973327
3	6	294	14.796175
4	7	178	8.958228

Informasi yang diperoleh:

1. Tiga customer teratas memiliki family members 3 sampai 5 orang
2. Pada histogram tersebut mayoritas customer memiliki 4 family members dengan presentase 25,41% dari keseluruhan data

2. Univariate Analysis

Univariate Analysis Data Kategorik – Histogram (1/3)

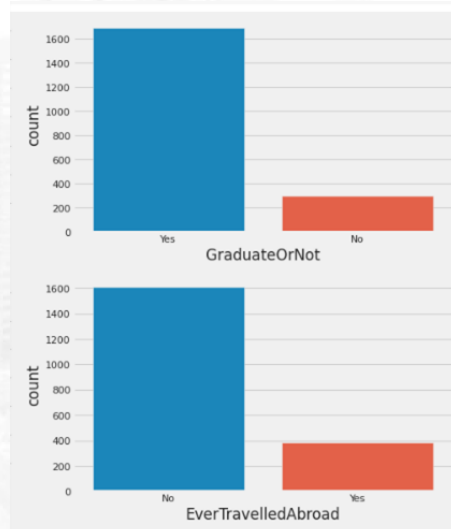


Informasi yang diperoleh:

1. Mayoritas customer dengan 71.31% dari keseluruhan populasi bekerja di private sector (perusahaan swasta) dan self employment (usaha mandiri)
2. Mayoritas customer yang terdata di dataset ini tidak mengidap penyakit kritis yaitu sekitar 72.22%

2. Univariate Analysis

Univariate Analysis Data Kategorik – Histogram (2/3)



GraduateOrNot	count	percent
0	Yes	1692 85.153498
1	No	295 14.846502

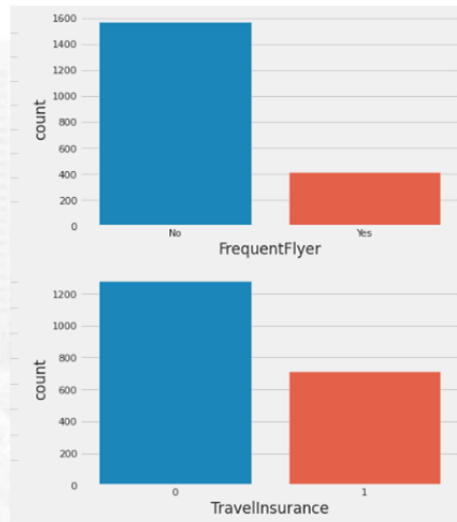
EverTravelledAbroad	count	percent
0	No	1607 80.875692
1	Yes	380 19.124308

Informasi yang diperoleh:

1. Mayoritas customer pada dataset ini merupakan lulusan kuliah (graduated) yaitu sekitar 85.15%
2. Sekitar 80.87% customer pada dataset ini belum pernah pergi ke luar negeri

2. Univariate Analysis

Univariate Analysis Data Kategorik – Histogram (3/3)



	FrequentFlyer	count	percent
0	No	1570	79.013588
1	Yes	417	20.986412

	TravelInsurance	count	percent
0	0	1277	64.26774
1	1	710	35.73226

Informasi yang diperoleh:

1. Sekitar 79.01% customer pada dataset ini belum pernah melakukan penerbangan.
2. Mayoritas customer yang terdata pada dataset ini belum pernah membeli asuransi perjalanan.
3. Berdasarkan data persentase variabel *Travel Insurance* kita mendapati bahwa data-nya imbalance dengan tingkatan ringan. Oleh karena itu, nantinya datanya bisa diproses untuk mengurangi imbalance-nya.

2. Univariate Analysis

Follow Up yang perlu dilakukan di Data Processing

Data Cleansing

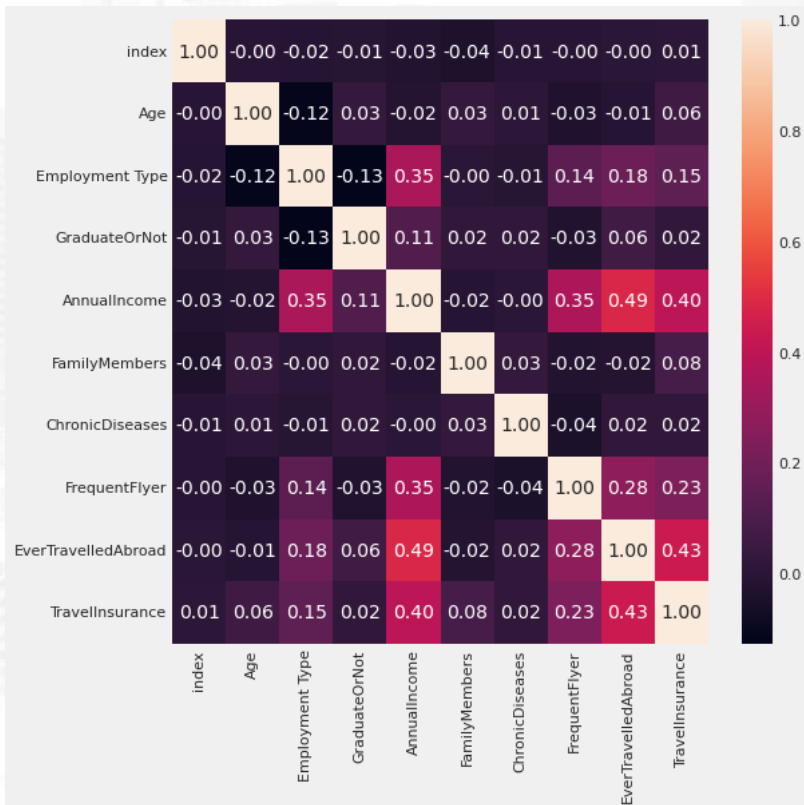
- Tidak perlu dilakukan handling missing value karena tidak adanya data yang null
- Terdapat 738 duplicate data yang perlu diremove.
- Terdapat beberapa features yang bertipe jenis kategorikal yang perlu dilakukan feature encoding dan transformation (FamilyMembers, FrequentFlyer, EverTravelledAbroad, ChronicDisseases, GraduateOrNot, dan Employment Type).
- Terpantau datasetnya imbalance dilihat dari persebaran kolom target travel insurance. Pada data processing perlu dilakukan Smote / Weight Class Sampling / Oversampling / Undersampling.
- Tidak terdapat adanya outliers sehingga tidak perlu dilakukannya handling outliers.

Feature Engineering

- Melakukan feature selection dengan menggunakan Chi Square (Feature Kategorik -> Target Kategorik) dan Anova (Feature Numerik -> Target Kategorik).
- Mungkin bisa melakukan feature extraction dengan membuat feature baru yakni income per capita yang didapat dari kolom (Annual Income / Family Member) dan lain sebagainya.

3. Multivariate Analysis

Heatmap



- Variabel Employment type, Annual Income, Frequent Flyer dan Ever Travelled Abroad Menunjukkan korelasi positif yang paling besar terhadap Travel Insurance daripada fitur lainnya.
- Keempat fitur di yang disebutkan sebelumnya kemungkinan akan dijadikan sebagai fitur utama sebab korelasinya terhadap fitur target yang cukup tinggi dibandingkan dengan fitur lainnya.
- Fitur AnnualIncome memiliki korelasi yang besar terhadap employment type, frequent flyer, serta evertravelledabroad yang mana fitur ini akan dilakukan pengecekan lebih lanjut apakah fitur-fitur tadi memiliki pengaruh satu sama lain terhadap pembelian TravelInsurance melalui penggunaan multi-bar plot.

3. Multivariate Analysis

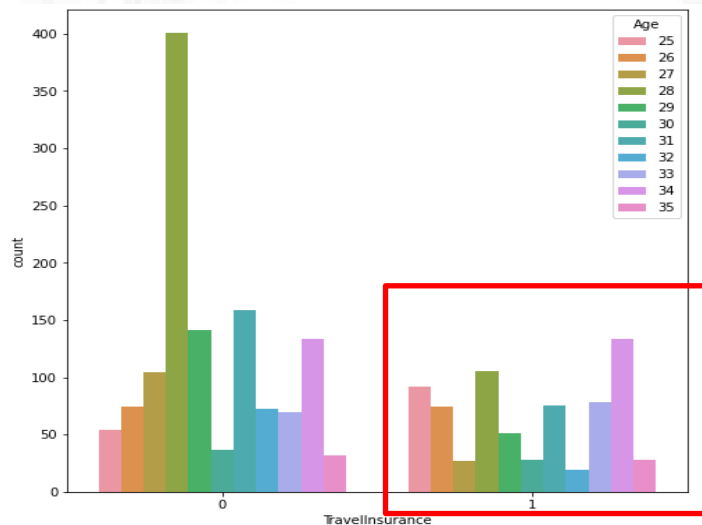
Pairplot



- Berdasarkan grafik pairplot diatas, kita dapat melihat bahwa jumlah data yang ada belum mampu menunjukkan korelasi dan distribusi dari keseluruhan data terhadap variabel lainnya.
- Perlu dilakukan pengecekan lebih lanjut dengan menggunakan multibarplot pada variabel terhadap feature vs Target sehingga dapat diperoleh gambaran besar untuk *business insight*.

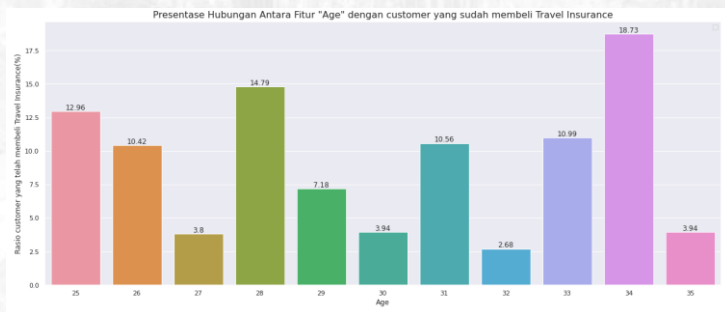
3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (1/8) Age vs TravelInsurance



Informasi yang diperoleh:

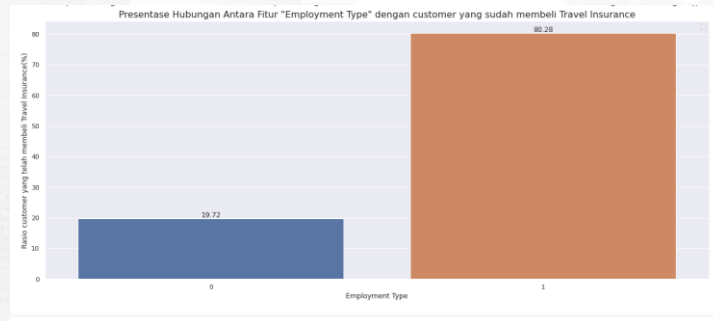
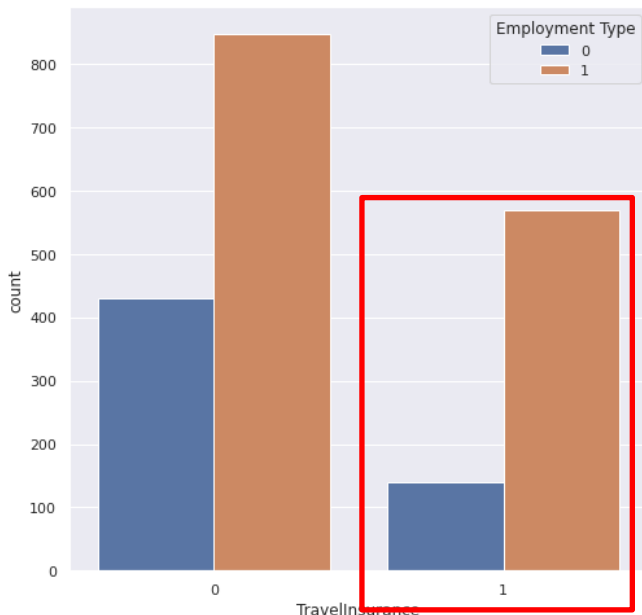
1. Mayoritas customer yang membeli travel insurance berada pada umur 34 tahun dengan rasio 18.73% dari list customer yang terdaftar.
2. Customer dengan umur 28 tahun merupakan customer mayoritas yang belum membeli travel insurance. Walaupun begitu customer dengan usia 28 tahun merupakan customer terbanyak pada dataset. Artinya, customer berusia 28 tahun dapat menjadi target dengan jumlah besar karena populasinya yang besar dan belum terjamah, namun memiliki ketertarikan pada travel insurance yang cukup tinggi.
3. Top 3 jumlah terbanyak yang membeli travel insurance ada pada umur 34, 28, dan 25.



3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (2/8)

EmploymentType vs TravelInsurance

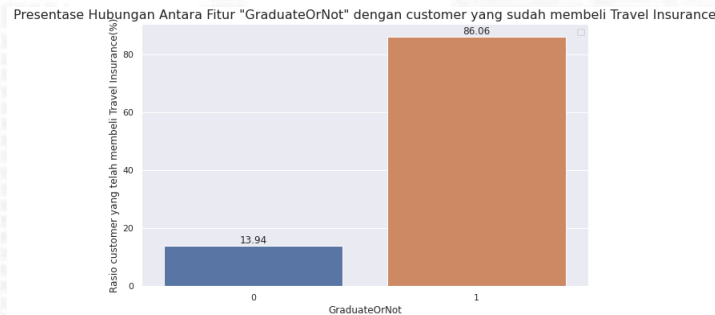
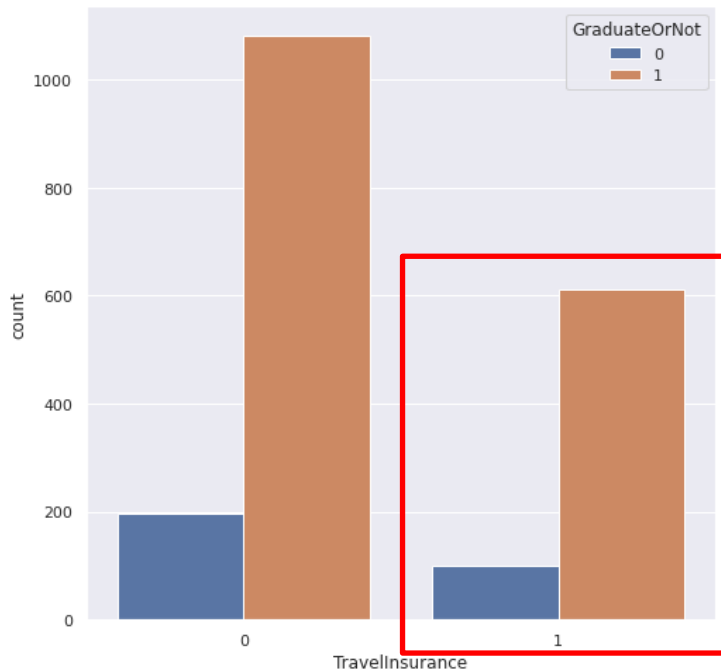


Informasi yang diperoleh:

1. Mayoritas customer yang membeli travel insurance bekerja di private sector/Self Employed dengan persentase 80.28%. Hal ini berbanding terbalik dengan customer yang bekerja di pemerintahan atau government sector yang hanya 19,72% saja. Penyebab ketimpangan ini mungkin dikarenakan di customer yang bekerja di perusahaan swasta atau wirausaha cenderung lebih sering berpergian ketimbang yang bekerja di pemerintahan.
2. Pada marketing campaign nantinya lebih baik difokuskan pada customer yang bekerja di perusahaan swasta atau yang bekerja sebagai wiraswasta.

3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (3/8) GraduateOrNot vs TravelInsurance

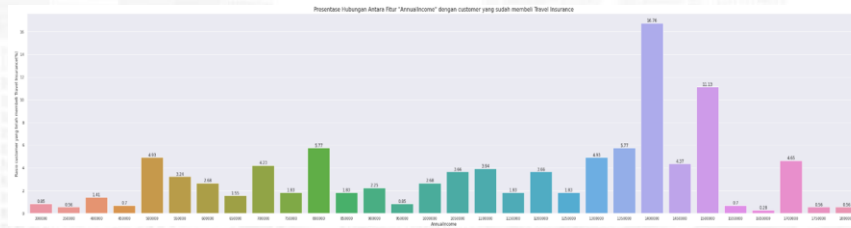
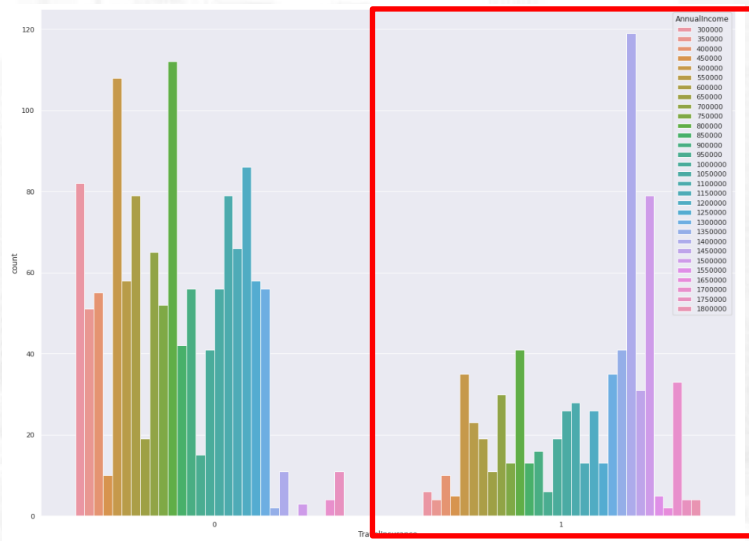


Informasi yang diperoleh:

1. Mayoritas customer yang membeli travel insurance belum lulus kuliah dengan persentase 86.06% dari list customer yang terdaftar di travel insurance.
2. Pada marketing campaign nantinya tidak perlu spesifik menargetkan lulusan kuliah atau saja, namun bisa lebih universal.

3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (4/8) AnnualIncome vs TravelInsurance

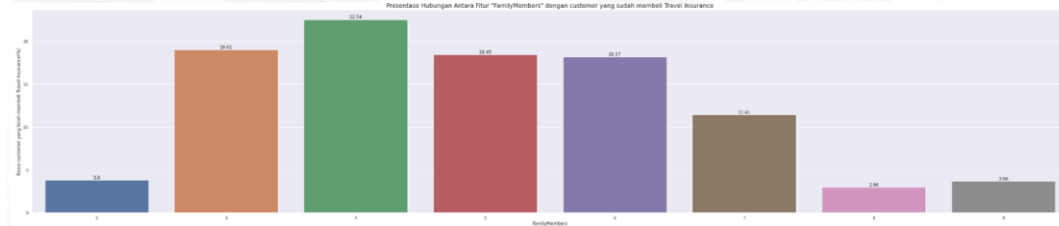
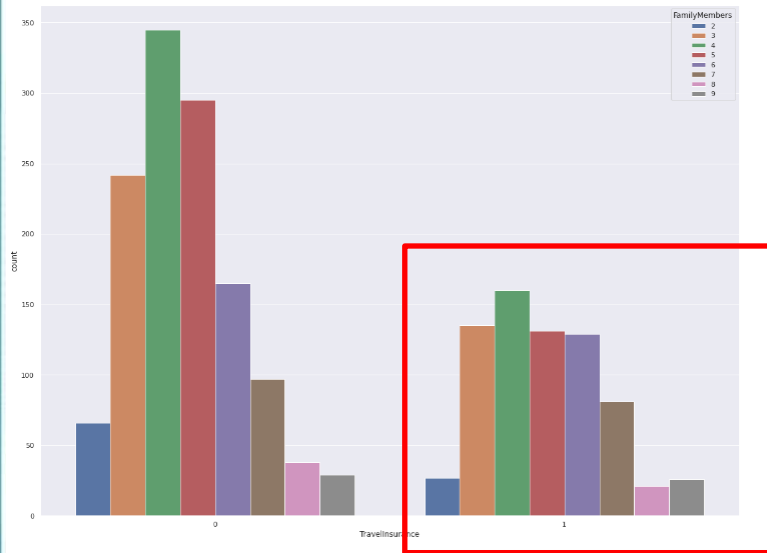


Informasi yang diperoleh:

1. Mayoritas customer yang membeli travel insurance berdasarkan grafik memiliki pendapatan yang lebih banyak ketimbang yang tidak.
2. Mayoritas customer yang memiliki travel insurance memiliki pendapatan 1,4 juta dengan 16,76%.
3. Pada saat marketing sebaiknya terlebih dahulu melakukan profiling atas pendapatan yang dimiliki customer dan memprioritaskan pendapatan yang besar terlebih dahulu.

3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (5/8) FamilyMember vs TravelInsurance

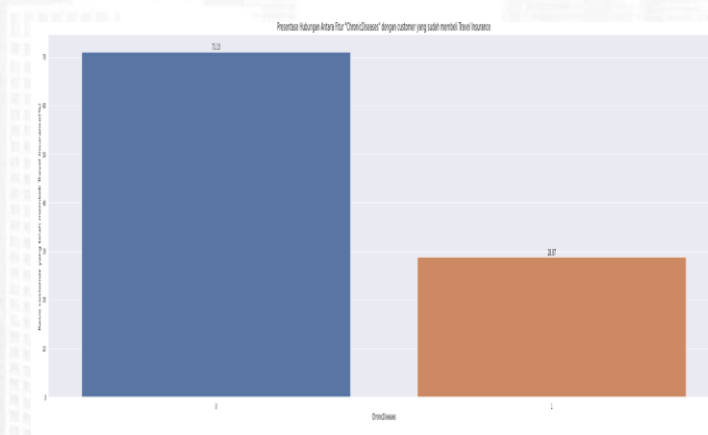
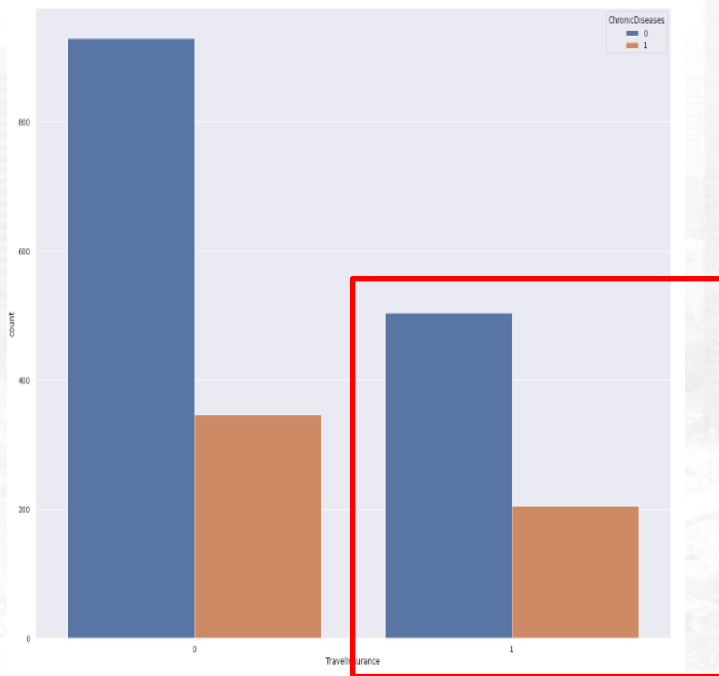


Informasi yang diperoleh:

1. Mayoritas customer yang membeli travel insurance memiliki jumlah anggota keluarga sebanyak 4 orang dengan presentase 22.54%.
2. Dapat disimpulkan bahwa jumlah anggota keluarga tidak terlalu berpengaruh terhadap travel insurance karena rasionya hampir sama. Namun, keluarga yang beranggotakan 8 hingga 9 orang tidak membeli banyak travel insurance dibandingkan keluarga lainnya

3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (6/8) ChronicalDisease vs TravelInsurance

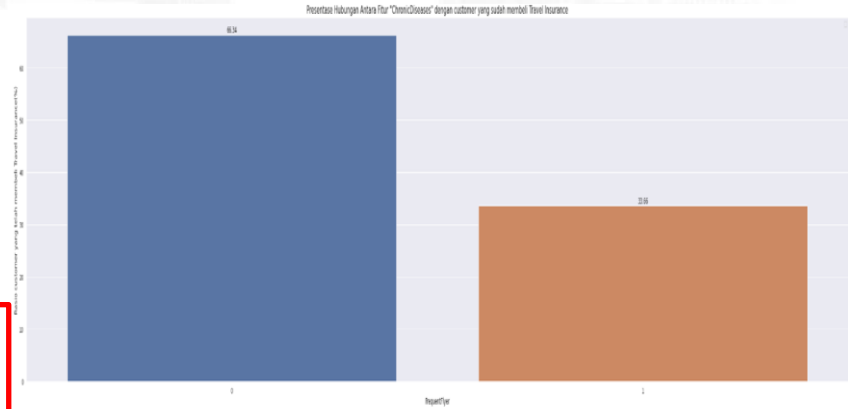
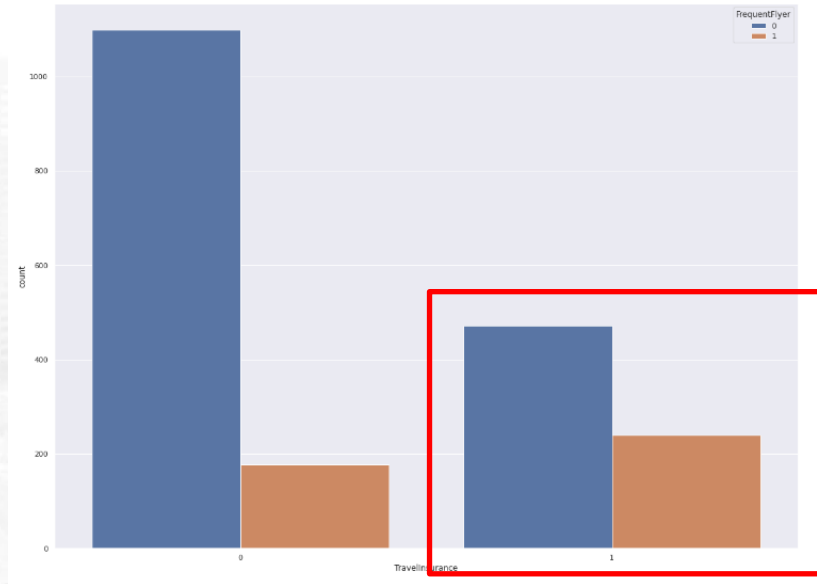


Informasi yang diperoleh:

1. Mayoritas customer yang membeli travel insurance tidak memiliki penyakit kronis dengan rasio 71.13%.
2. Dari grafik di atas kita dapat berasumsi bahwa penyakit kurang berdampak pada pembelian asuransi perjalanan atau tidak.

3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (7/8) FrequentFlyer vs TravelInsurance

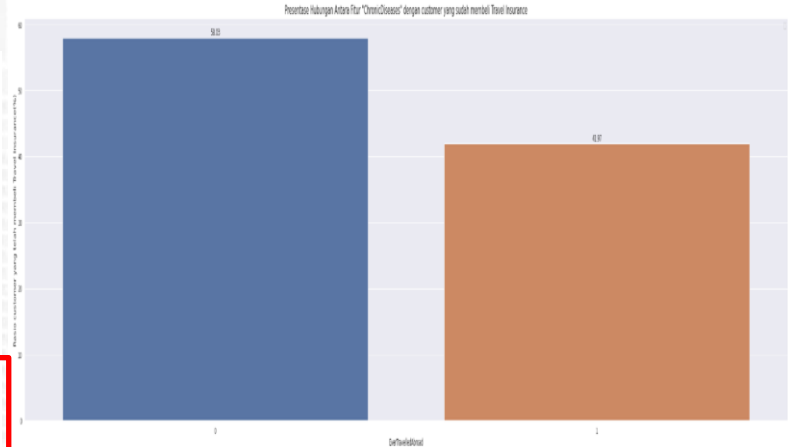
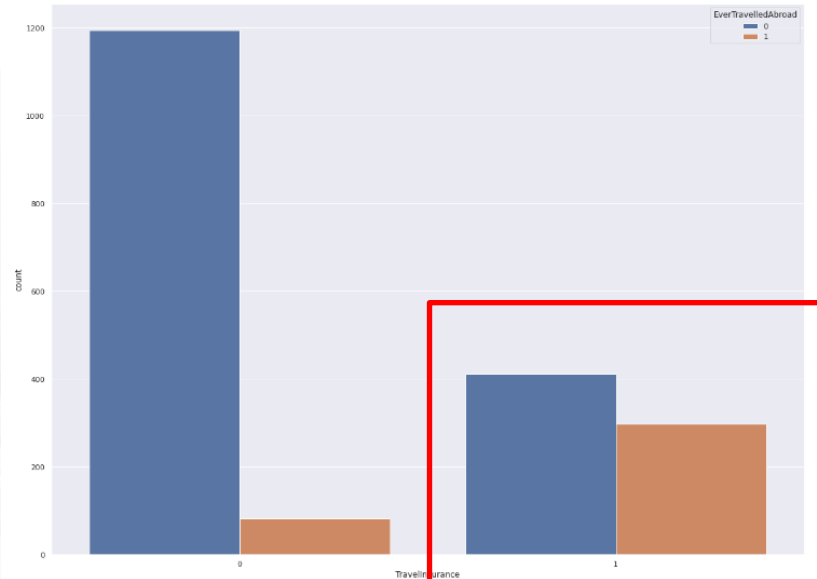


Informasi yang diperoleh:

1. Customer yang tidak sering melakukan penerbangan memiliki rasio untuk membeli travel insurance lebih besar dengan rasio sekitar 66.34%.
2. Customer yang sering travel menggunakan pesawat lebih cenderung membeli asuransi perjalanan sedangkan ada sebagian kecil customer yang tidak sering travel menggunakan pesawat tetapi masih memiliki travel insurance.

3. Multivariate Analysis

Multivariate Analysis Feature vs Target – Pair Plot (8/8) EverTravelAbroad vs TravelInsurance

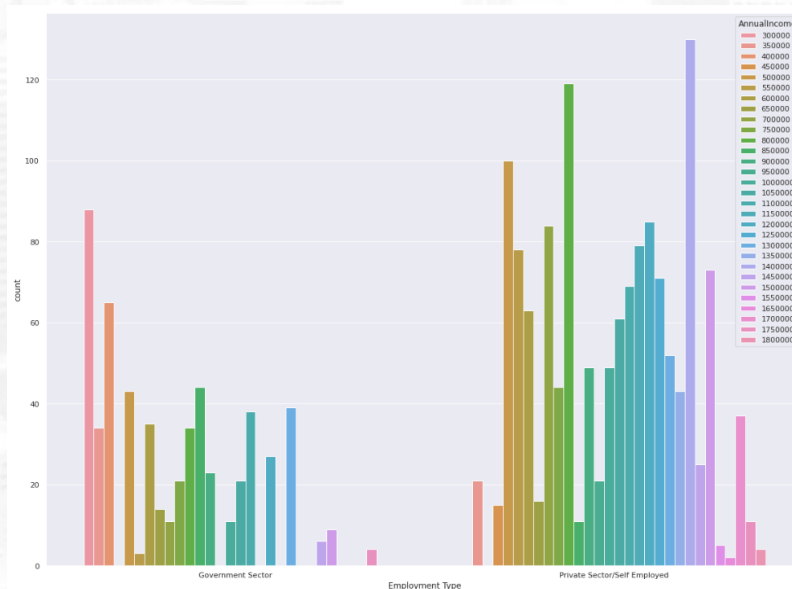


Informasi yang diperoleh:

1. Mayoritas customer yang membeli travel insurance belum pernah pergi ke luar negeri dengan rasio 58.03%.
2. Seperti yang kita ketahui, bahwa mayoritas masyarakat yang belum bepergian ke luar negeri dan belum membeli asuransi. Jadi satu hal yang dapat kita asumsikan bahwa orang yang belum bepergian ke luar negeri tidak memerlukan travel insurance saat bepergian di dalam negara yang sama.

3. Multivariate Analysis

Multivariate Analysis Feature vs Feature – Pair Plot (1/5) AnnualIncome vs EmploymentType

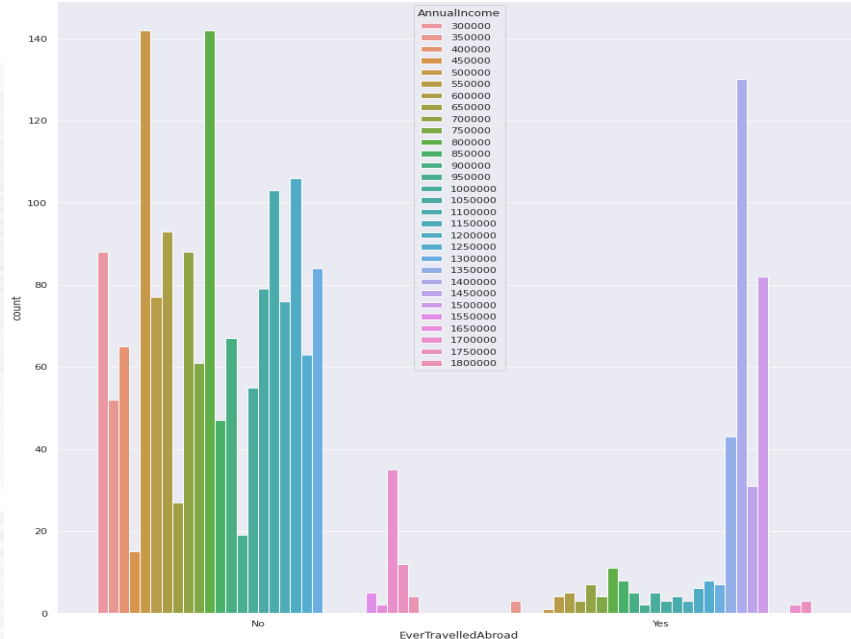


Informasi yang diperoleh:

Berdasarkan grafik dapat diambil kesimpulan bahwa customer yang bekerja di private sector (perusahaan swasta) /self employed (wiraswasta) cenderung memiliki penghasilan yang lebih besar ketimbang yang bekerja di pemerintahan atau government sector.

3. Multivariate Analysis

Multivariate Analysis Feature vs Feature – Pair Plot (2/5) AnnualIncome VS EverTravelledAbroad

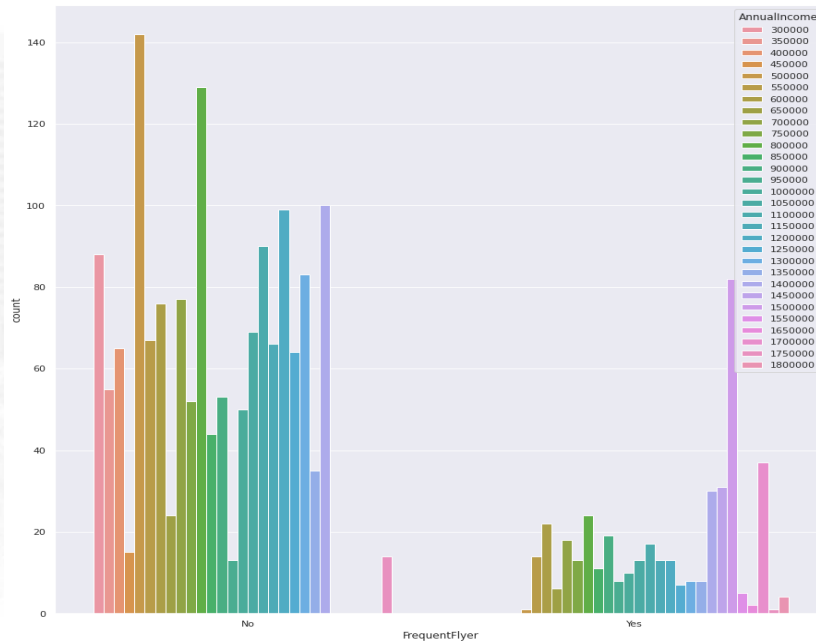


Informasi yang diperoleh:

1. Berdasarkan grafik dapat diambil kesimpulan bahwa jumlah penghasilan tidak terlalu menentukan apakah customer pernah travel keluar negeri atau tidak. Walaupun begitu, dapat terlihat bahwa masyarakat dengan penghasilan dibawah 1.400.000 rupee cenderung tidak pernah travel ke luar negeri.

3. Multivariate Analysis

Multivariate Analysis Feature vs Feature – Pair Plot (3/5) AnnualIncome vs FrequentFlyer

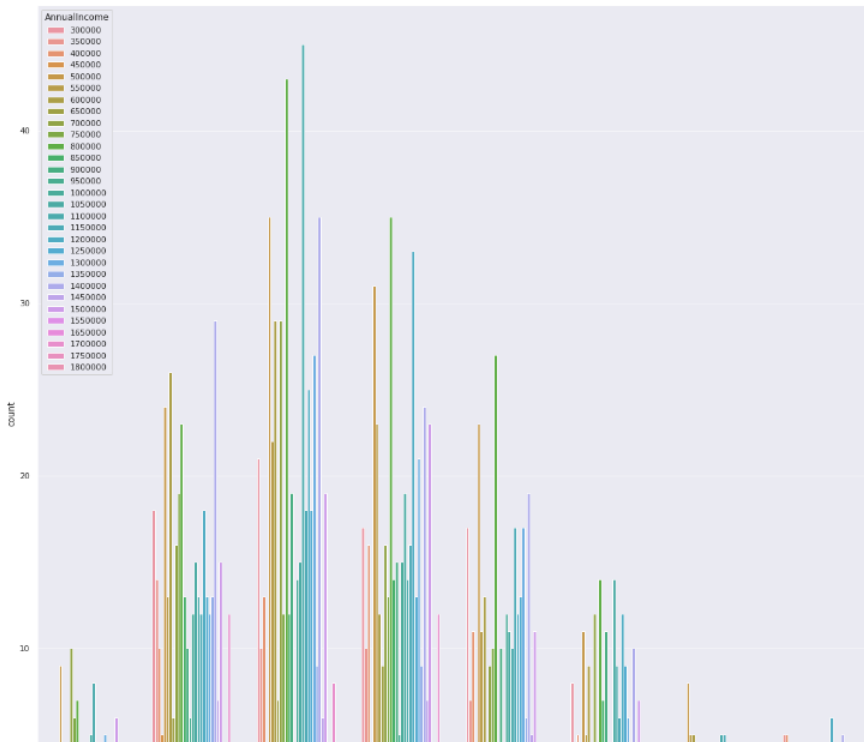


Informasi yang diperoleh:

1. Berdasarkan grafik dapat diambil kesimpulan bahwa jumlah penghasilan juga tidak terlalu menentukan apakah customer akan sering travel menggunakan pesawat atau tidak. Walaupun begitu dapat terlihat bahwa masyarakat dengan penghasilan dibawah 450.000 rupee cenderung tidak sering travel menggunakan pesawat.

3. Multivariate Analysis

Multivariate Analysis Feature vs Feature – Pair Plot (4/5) AnnualIncome vs FamilyMember

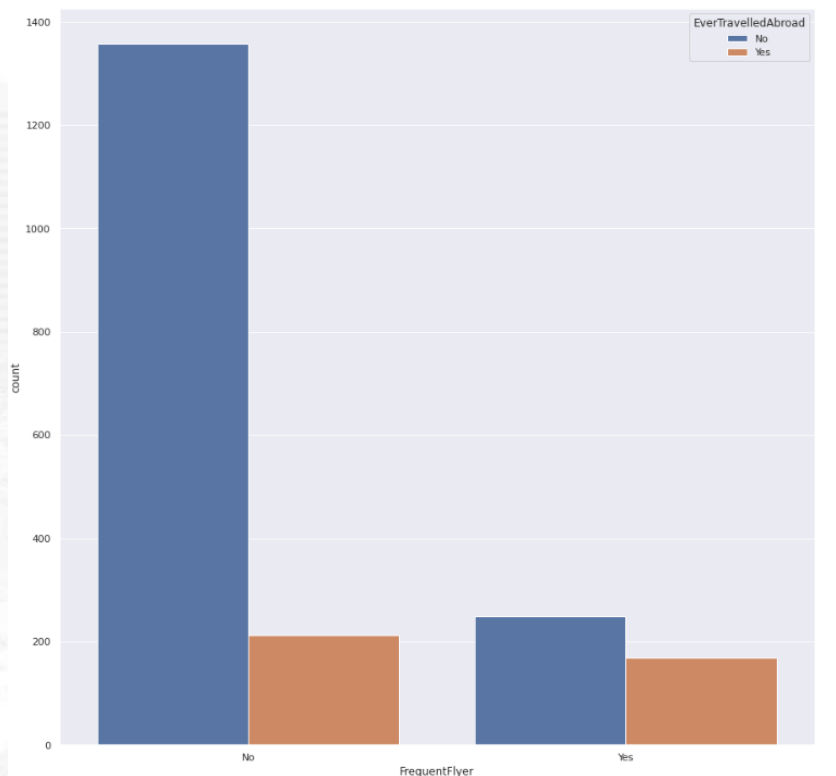


Informasi yang diperoleh:

Berdasarkan grafik dapat terlihat bahwa customer dengan penghasilan terbesar atau diatas 1,3 juta mayoritas memiliki 4 anggota keluarga. Berdasarkan hal ini mungkin bisa dilihat income/capita tiap customer.

3. Multivariate Analysis

Multivariate Analysis Feature vs Feature – Pair Plot (5/5) EverTravelAbroad vs FrequentFlyer



Informasi yang diperoleh:

Berdasarkan grafik dapat dilihat bahwa orang yang sering menggunakan bepergian menggunakan pesawat hampir setengahnya pernah berpergian menggunakan pesawat. Walaupun begitu, mayoritas customer yang sering berpergian menggunakan pesawat belum pernah ke luar negeri.

3. Multivariate Analysis

A. Korelasi Setiap Feature dan Label

Usulan feature yang perlu dipertahankan

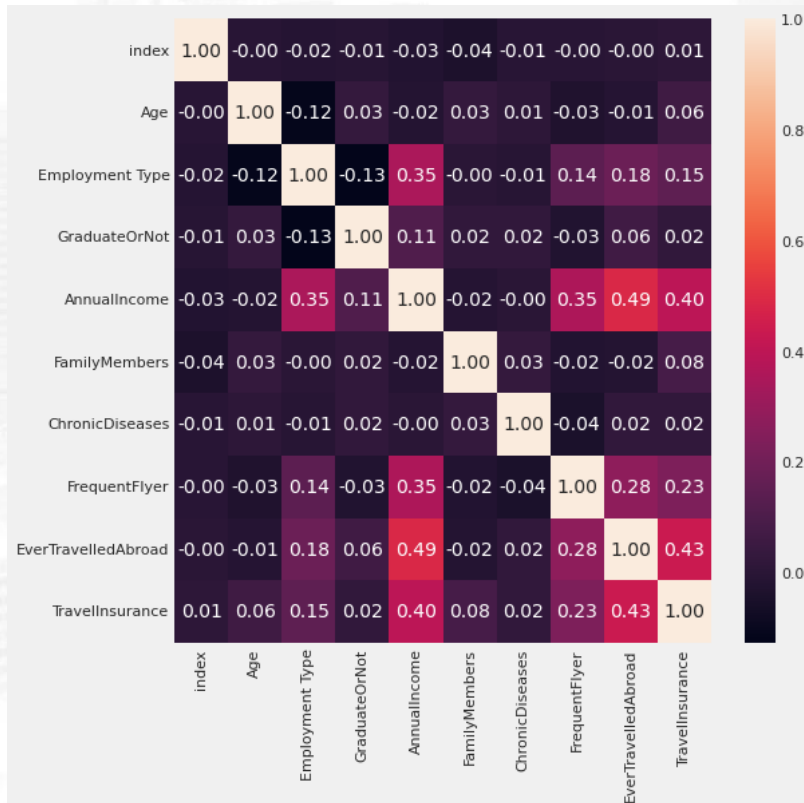
- Annual Income
- Ever Travel Abroad
- Frequent Flyer
- Employment Type
- Family Member
- Age

B. Korelasi Antar Feature

- Tidak ada feature yang redundan
- Mungkin bisa dilakukan feature extraction antara family member dan annual income karena menjadi sumber insight baru

4. Business Insight

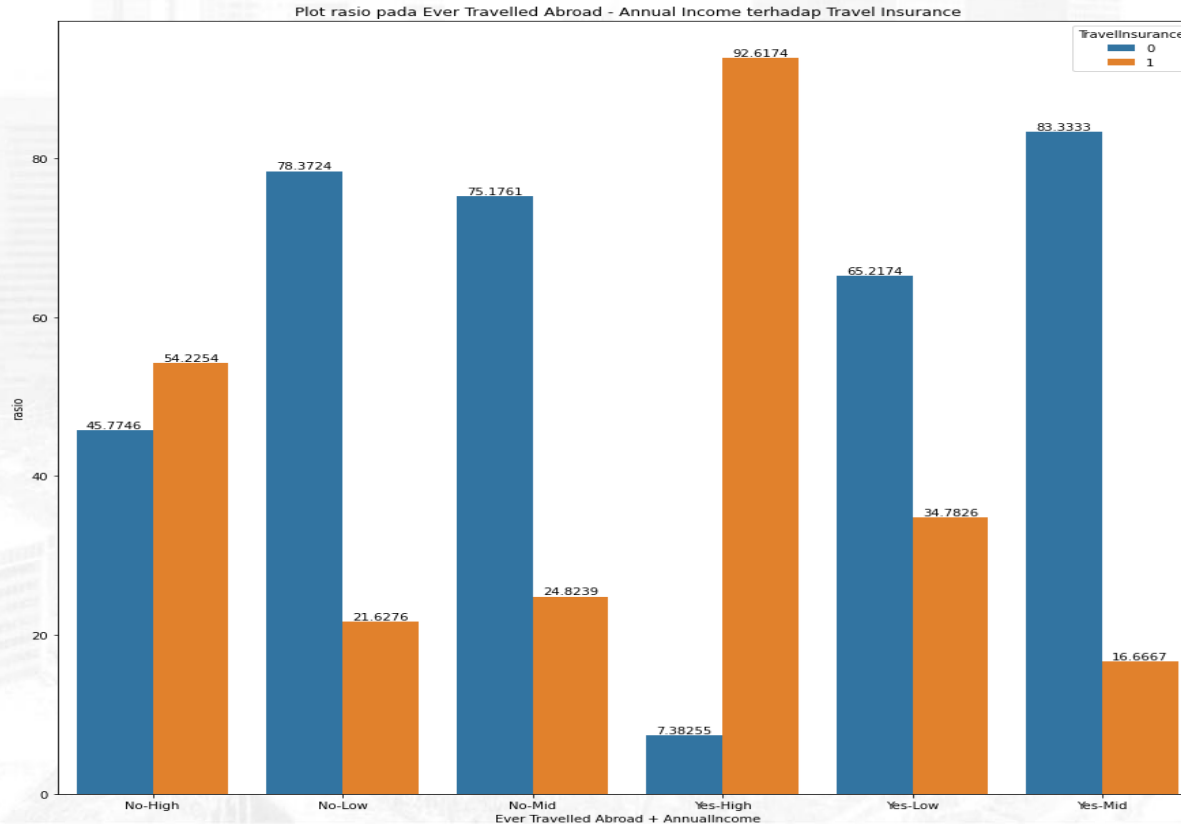
Heatmap



- Beberapa fitur yang memiliki korelasi positif terhadap satu sama lainnya seperti annual income - employment type, annual income - frequent flyer, annual income - ever travelled abroad, frequent flyer - ever travelled abroad.
- Beberapa fitur yang memiliki korelasi positif akan dilakukan pengecekan lebih lanjut agar diketahui lebih spesifik hubungannya satu sama lain dan apakah bisa berdampak pada keputusan membeli travel insurance ke

4. Business Insight

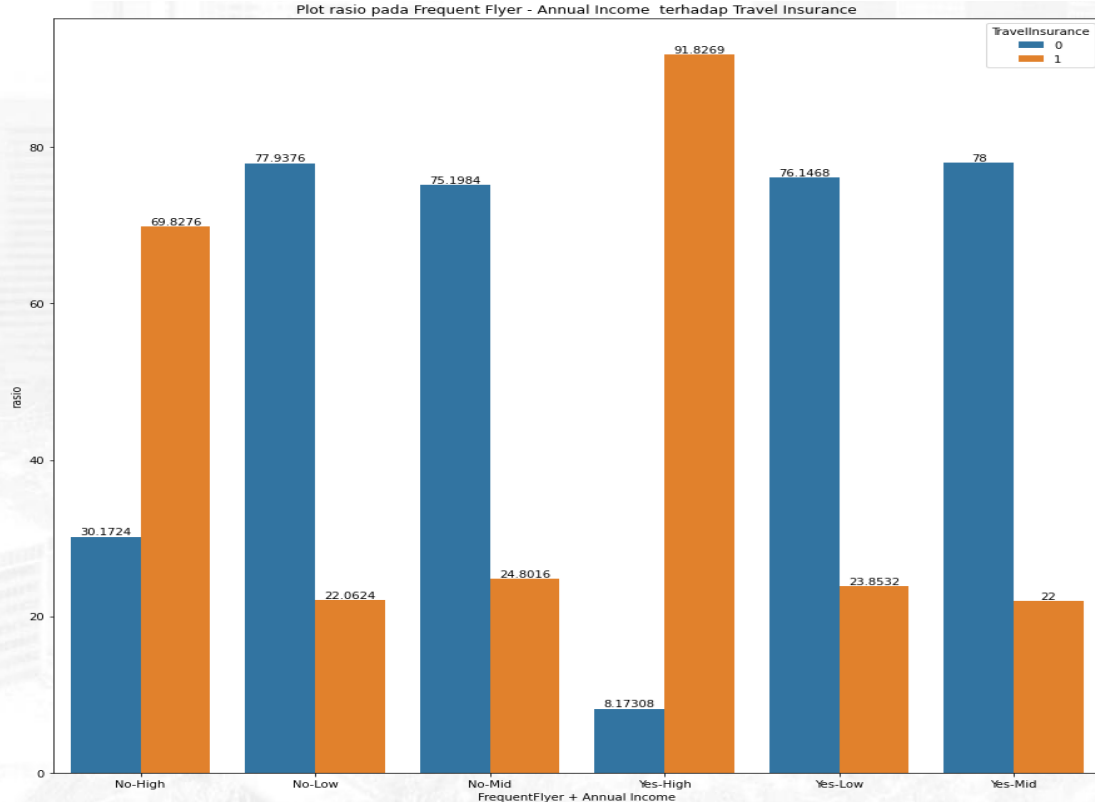
Plot rasio pada EverTravelled Abroad dan Annual Income terhadap Travel Insurance



- Dengan menggunakan multi barplot dari fitur yang dipilih menggunakan heatmap, kita mendapatkan gambaran bahwasanya:4
- Orang-orang yang pernah berpergian ke luar negeri dan memiliki penghasilan yang tinggi(setelah dilakukan segmentasi) memiliki kecenderungan yang paling tinggi untuk membeli travel insurance, diikuti dengan mereka yang tidak pernah ke luar negeri dan memiliki penghasilan tinggi.

4. Business Insight

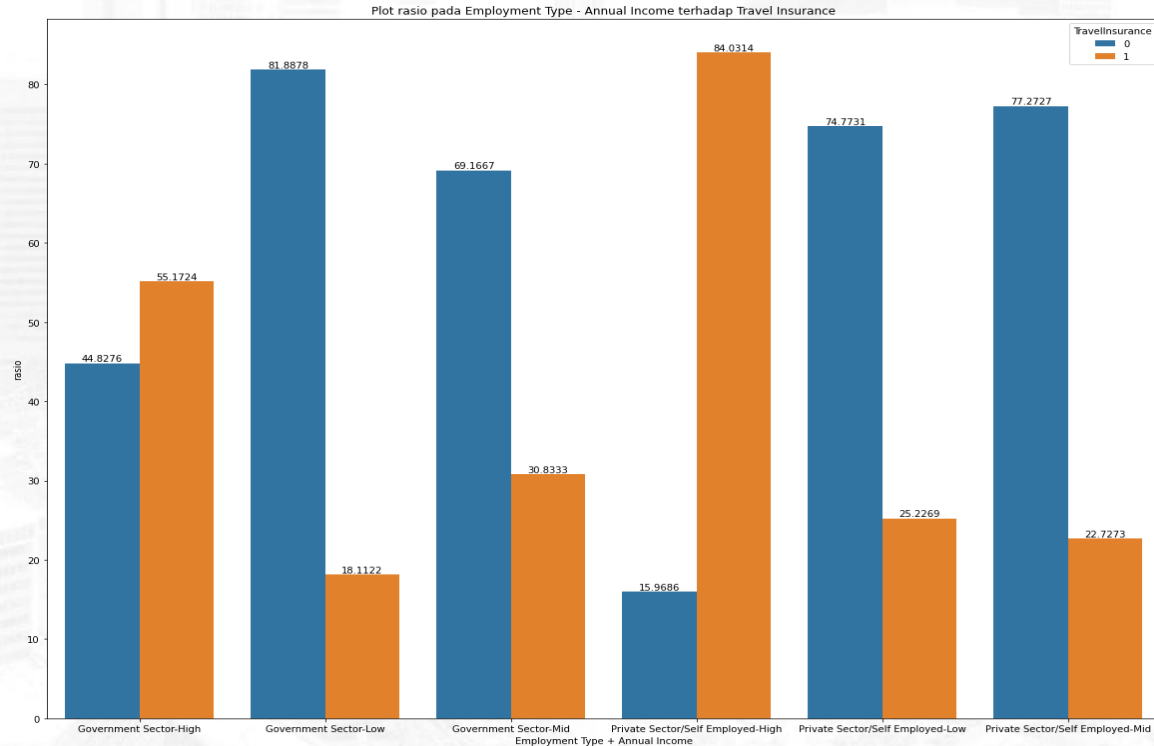
Plot rasio pada Frequent Flyer dan Annual Income terhadap Travel Insurance



- Dari plot di samping dapat dilihat bahwasanya orang-orang yang sering terbang dan berasal dari segment pendapatan yang tinggi memiliki kecenderungan untuk membeli travel insurance dibanding mereka yang sering terbang tetapi berasal dari segment pendapatan low atau mid, bahkan yang berasal dari segment low punya kecenderungan untuk membeli travel insurance daripada orang-orang yang sering terbang dan berasal dari segment pendapatan mid.

4. Business Insight

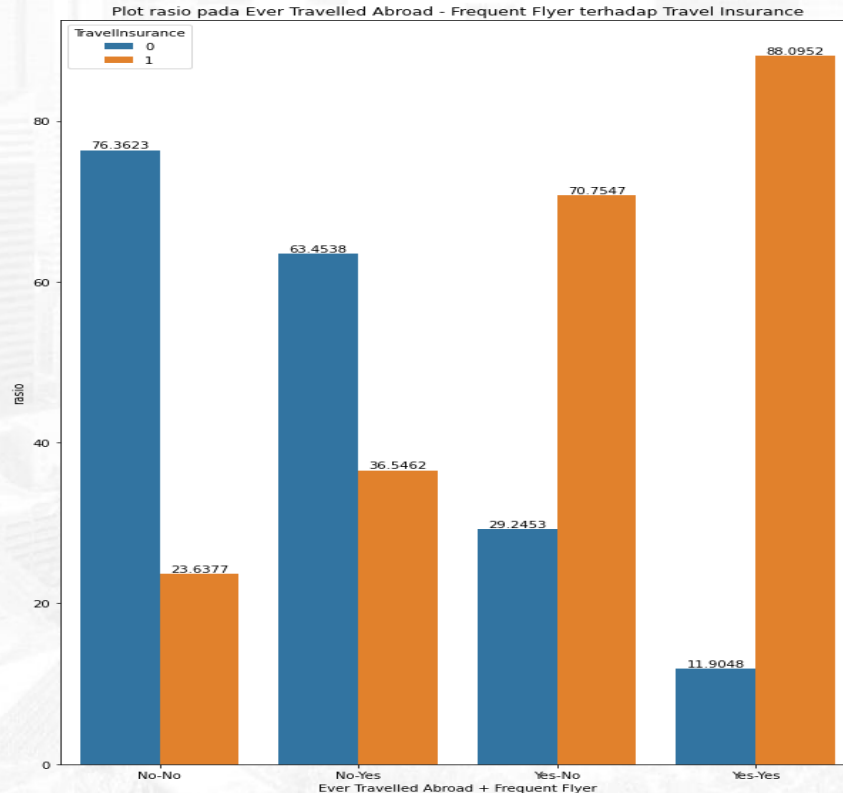
Plot rasio pada Employment Type dan Annual Income terhadap Travel Insurance



- Dari plot di samping juga dapat dilihat bahwasanya orang-orang dari private sector dan berasal dari segment pendapatan High memiliki rasio pembelian travel insurance lebih tinggi daripada orang-orang yang berasal dari sektor pemerintahan dan berasal dari segment pendapatan yang sama, ini juga mungkin disebabkan dari kecenderungan orang-orang dari private sector yang sering melakukan penerbangan.

4. Business Insight

Plot rasio pada Ever Travel Abroad dan Frequent Flyer terhadap Travel Insurance



- Juga terlihat dari plot di samping bahwasanya orang-orang yang pernah terbang ke luar negeri dan juga sering melakukan penerbangan memiliki nilai pembelian insurance yang paling tinggi, diikuti dengan mereka yang pernah terbang ke luar negeri akan tetapi jarang melakukan penerbangan.

4. Business Insight

Dari insight yang didapatkan melalui proses EDA tim data scientist sebagai data consultant di perusahaan UNO Inc. memberikan beberapa rekomendasi berupa:

1. Tim marketing dapat membuat campaign yang berfokus kepada mereka yang berasal dari orang-orang yang sering melakukan penerbangan, orang-orang yang pernah melakukan penerbangan ke luar negeri, memiliki penghasilan yang tinggi dan bekerja di private sector sehingga lebih mudah untuk menarik perhatian dan minat mereka untuk membeli travel insurance.
2. Tim marketing juga bisa mengurangi atau memberhentikan pemberian campaign kepada mereka yang berasal dari luar 4 kategori tadi agar cost campaign bisa menurun atau bisa dialokasikan kepada pemberian campaign kepada customer yang masuk ke dalam salah satu dari 4 kategori tadi.
3. Campaign yang diberikan dapat berupa awareness kepada para customer atau calon customer yang masuk ke dalam salah satu dari ke-4 kategori tadi agar nantinya diharapkan mereka akan lebih memahami fungsi dari travel insurance yang disediakan oleh UNO Inc.

5. Git

Link Repository GIT Group 1 UNO

The background of the slide is a faded, grayscale aerial photograph of a dense urban skyline, likely New York City, showing numerous skyscrapers and a complex network of roads and green spaces.

[Link Google Drive Group 1 UNO](#)