



Holiday Package Prediction

Final Project
Kelompok 3 Batch 32

OUR TEAM



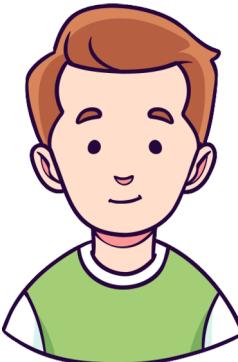
Arnanda

R. Arnanda Adi
Wijanarko



Edgar

Edgar Ariel
Majied



Teguh

Teguh
Ferdianto



Sendhy

Sendhy
Boedhi



Vionella

Vionella
Awanda
Irsabadi



Jannisah

Jannisah Dwi
Rahhadiski



Faris

Faris Isham
Wiryansyah



Jodhi

Jodhi
Krisantus
Sihalbu



OUTLINE

Stage 1

Business Understanding

Stage 2

Exploratory Data Analysis

Stage 3

Data Pre-Processing

Stage 4

Modeling & Evaluation

Stage 5

Business Simulation

Stage 6

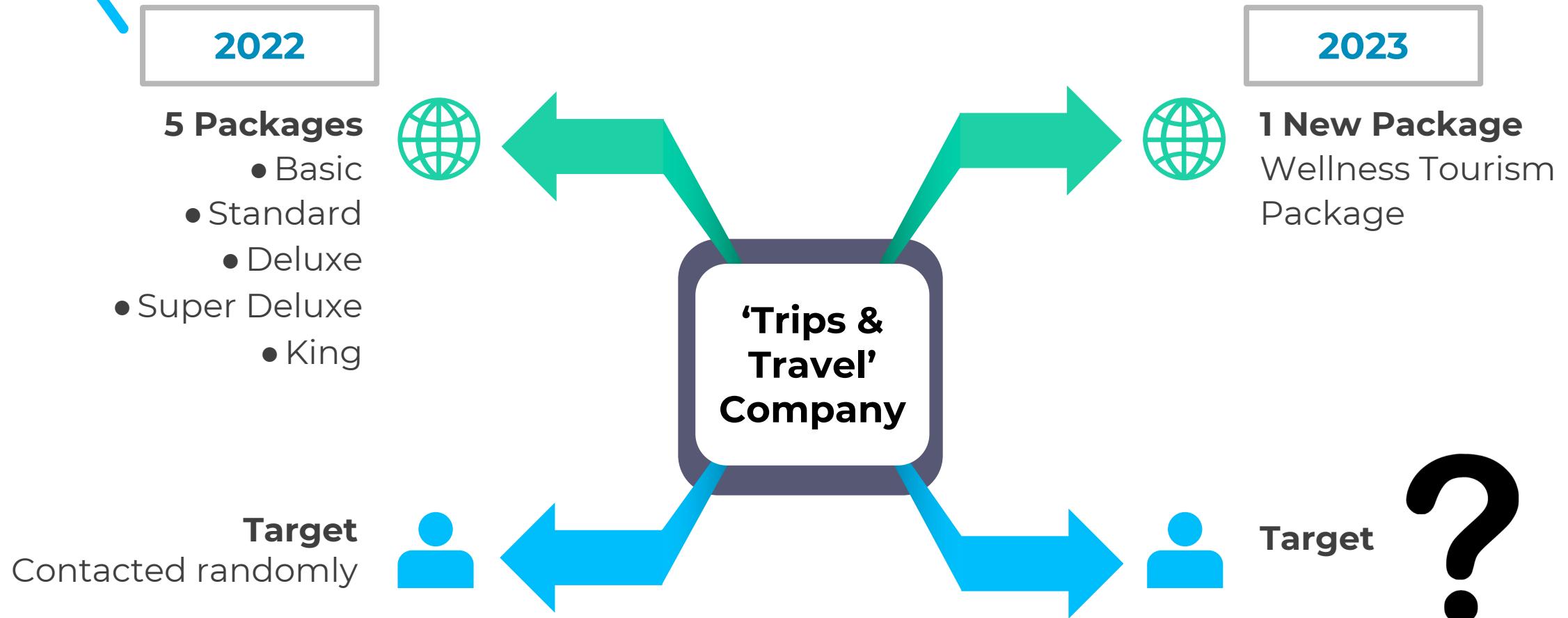
Business Recommendation



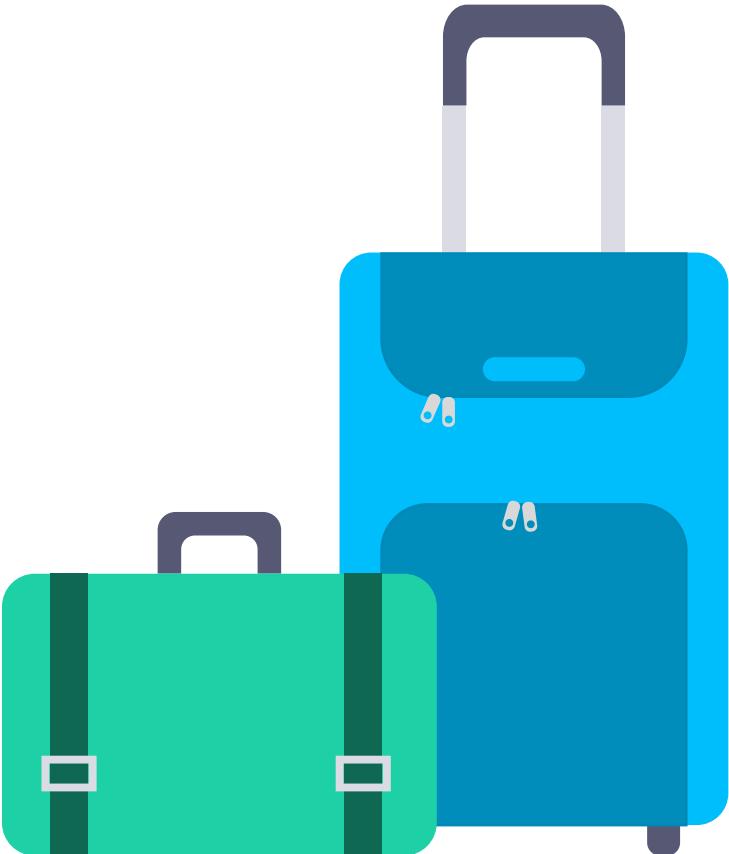
BUSINESS UNDERSTANDING



Problem Statement



- ❖ Tahun lalu **hanya 19%** pelanggan yang membeli paket liburan.
- ❖ **Revenue** perusahaan **tidak** mengalami peningkatan yang signifikan.



Goals

Untuk menaikkan conversion rate dan revenue perusahaan



Objectives

Membuat model untuk memprediksi pelanggan yang akan membeli paket liburan terbaru



Business Metrics

Total revenue dan conversion rate



EXPLORATORY DATA ANALYSIS



Dataset Analysis



Jumlah Baris

4888

Jumlah Kolom

20 kolom

Missing Value

8 kolom

Duplicated

No duplicate data

Target

Kolom **ProdTaken**

Kesalahan Value

Kolom *Gender* value
'Fe Male'

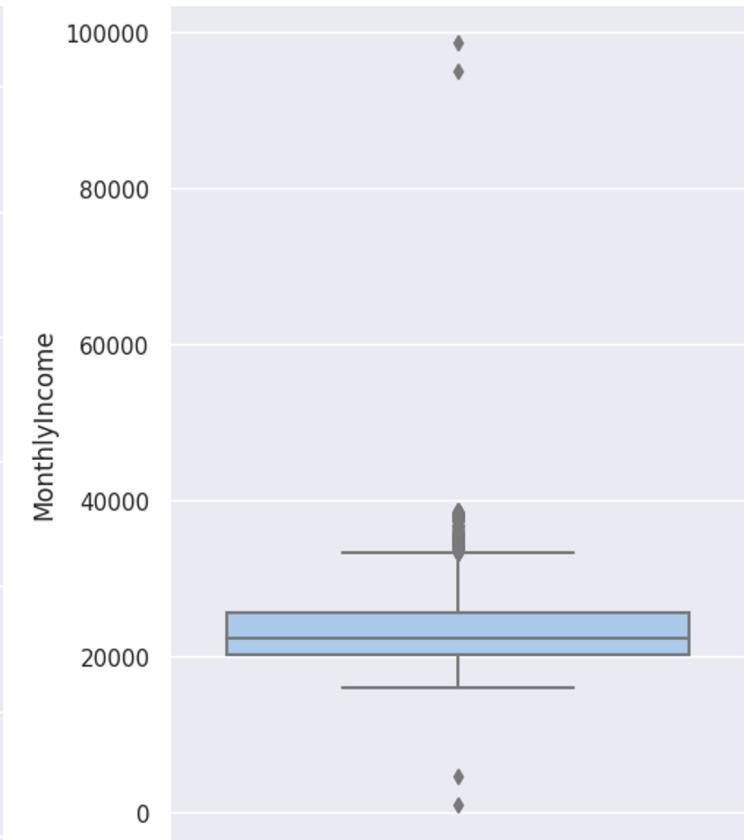
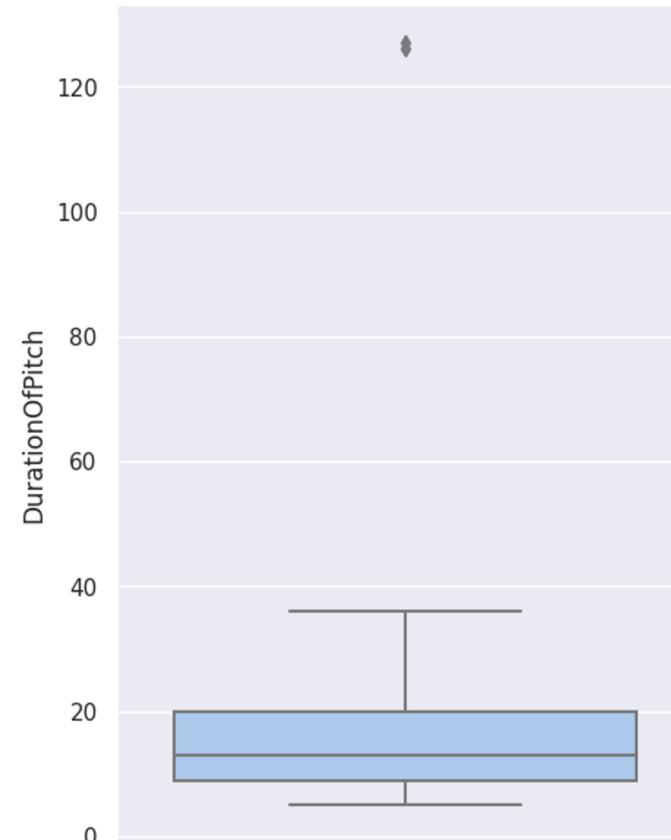
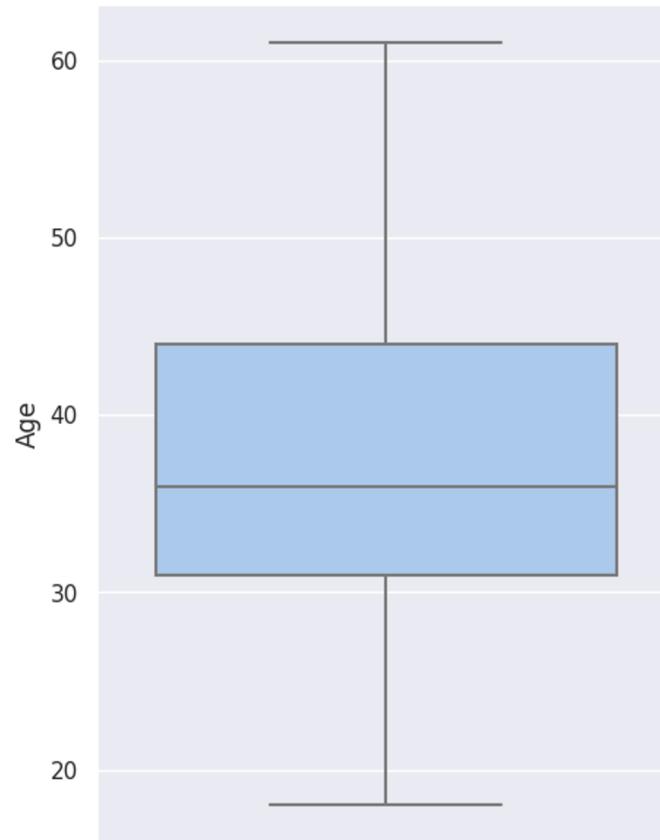
Kemiripan Value

Kolom *MaritalStatus*
value **'Divorced'**, **'Single'**,
dan **'Unmarried'**

Tipe Data Kurang Sesuai

7 kolom dengan tipe data
float, diubah menjadi tipe
data **integer**

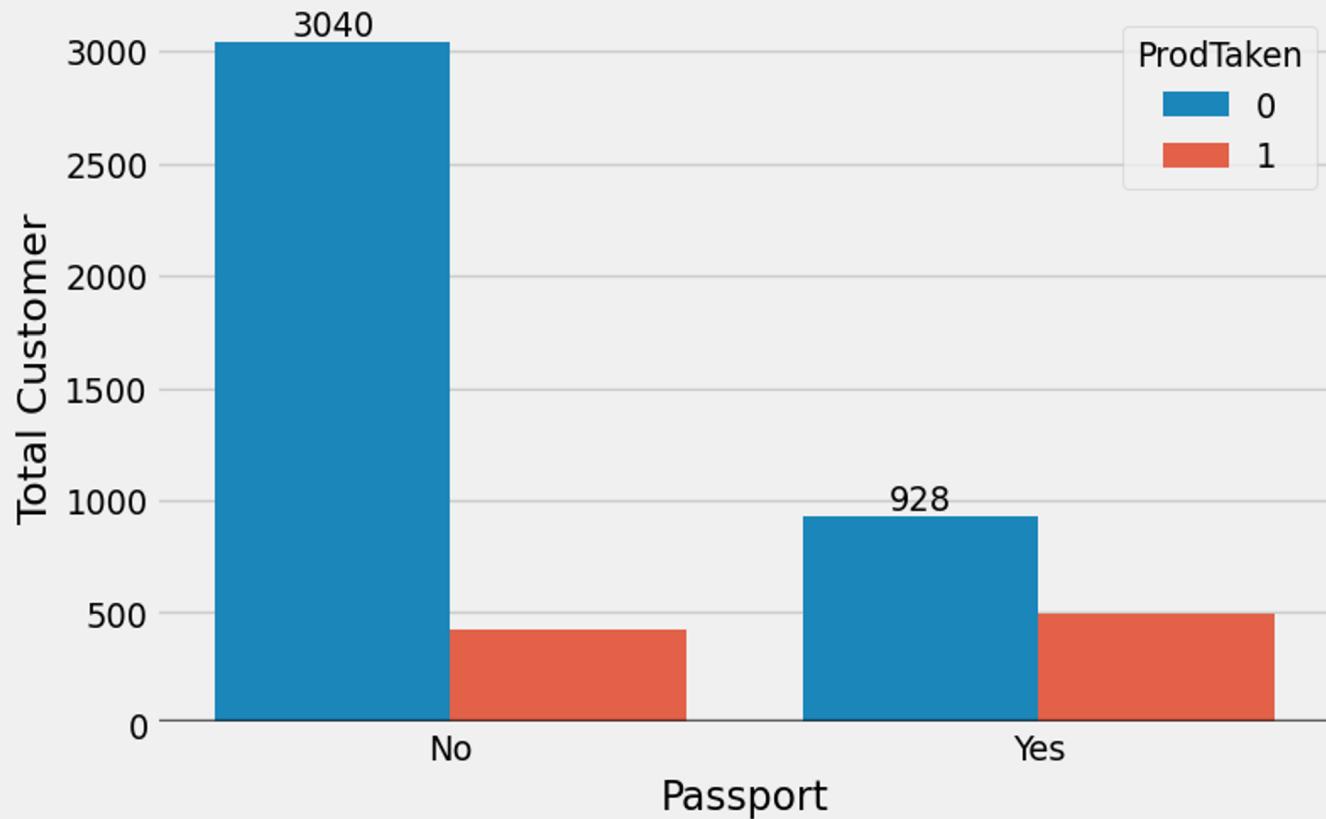
Dataset Analysis



Terdapat 2 kolom yang memiliki outliers (data dengan nilai yang menyimpang atau jauh dari pola).

Insights

Pelanggan yang Menolak Tawaran Paket Liburan Didominasi oleh Pelanggan yang Tidak Memiliki Passport

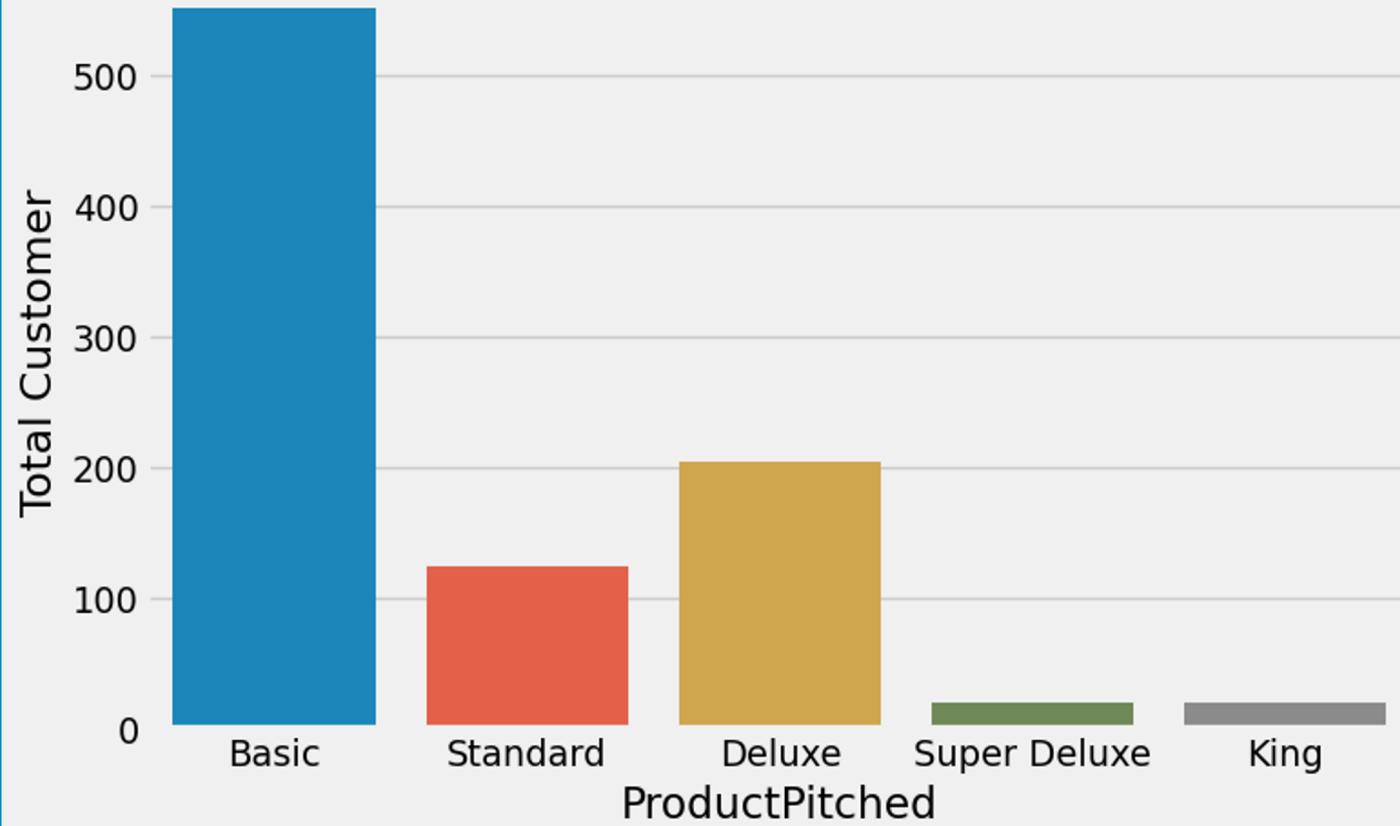


©Pandas Lovers : Holiday Package Travel



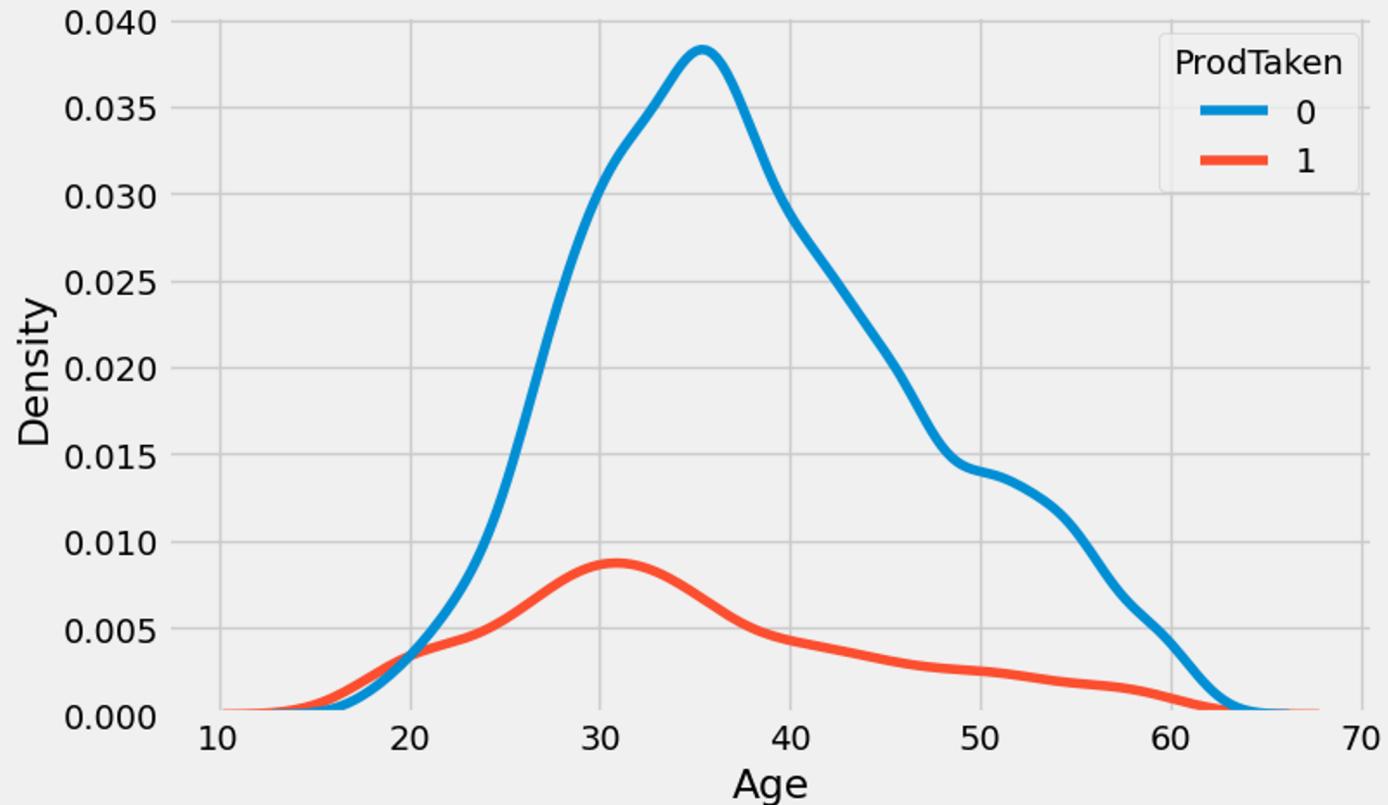
Insights

Pelanggan Cenderung Membeli Paket Kelas Rendah Hingga Paket Kelas Menengah (Basic - Deluxe)



Insights

**Pelanggan yang Berumur Muda ($\pm 20-35$ tahun)
Cenderung Lebih Membeli Tawaran Paket Liburan**



© Pandas Lovers : Holiday Package Travel



DATA PRE- PROCESSING



Data Cleansing

Handling Missing

Values

Mean

Age

Drop

TypeOfContact,

PreferredPropertyStar,

NumberOfTrips,

NumberOfChildrenVisiting

Median

DurationOfPitch, MonthlyIncome

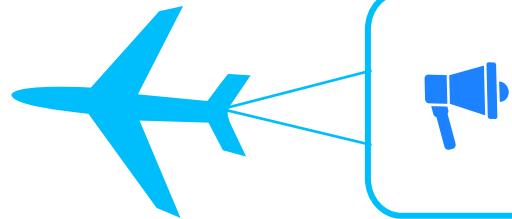
Modus

NumberOfFollowups



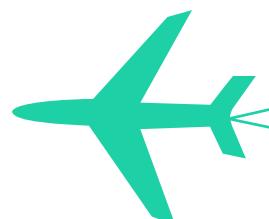
Handling Outliers

Tidak dilakukan karena model non-linear (model yang robust)



Handling Invalid Data Type

7 kolom diubah dari tipe data float menjadi integer



Handling Invalid Value

Kesalahan penulisan 'Fe Male'
diubah menjadi '**Female**'

Feature Transformation



Kolom Age:
standarisasi



Kolom *DurationOfPitch*
dan *MonthlyIncome*:
normalisasi

Feature Encoding

- **One-hot encoding** pada kolom *Occupation* dan *TypeOfContact*
- **Label encoding** pada kolom *Gender*, *ProductPitched*, *MaritalStatus*, dan *Designation*

Split Data

- **70%** data train
- **30%** data test



Correlation

Untuk melihat apakah di antara fitur yang tersisa, terdapat fitur yang redundant atau tidak



	ProdTaken -	Age -	CityTier -	DurationOfPitch -	NumberOfFollowups -	ProductPitched -	PreferredPropertyStar -	NumberOfTrips -	Passport -	PitchSatisfactionScore -	MonthlyIncome -	
ProdTaken -	1.00	-0.14	0.08	0.08	0.12	-0.21	0.10	0.02	0.26	0.06	-0.13	1.00
Age -	-0.14	1.00	-0.01	-0.00	-0.03	0.36	-0.02	0.17	0.03	0.01	0.41	
CityTier -	0.08	-0.01	1.00	0.02	0.03	0.22	-0.01	-0.03	0.00	-0.04	0.09	
DurationOfPitch -	0.08	-0.00	0.02	1.00	0.02	0.01	-0.01	0.01	0.03	-0.01	0.03	
NumberOfFollowups -	0.12	-0.03	0.03	0.02	1.00	-0.04	-0.02	0.14	0.01	-0.00	0.16	
ProductPitched -	-0.21	0.36	0.22	0.01	-0.04	1.00	-0.03	0.04	-0.02	0.01	0.57	
PreferredPropertyStar -	0.10	-0.02	-0.01	-0.01	-0.02	-0.03	1.00	0.01	0.00	-0.02	0.01	
NumberOfTrips -	0.02	0.17	-0.03	0.01	0.14	0.04	0.01	1.00	0.01	-0.00	0.13	
Passport -	0.26	0.03	0.00	0.03	0.01	-0.02	0.00	0.01	1.00	0.01	0.01	
PitchSatisfactionScore -	0.06	0.01	-0.04	-0.01	-0.00	0.01	-0.02	-0.00	0.01	1.00	0.02	
MonthlyIncome -	-0.13	0.41	0.09	0.03	0.16	0.57	0.01	0.13	0.01	0.02	1.00	

MODELING & EVALUATION



Modeling

Penggunaan precision karena ingin menurunkan angka false positive (FP).
False positive adalah kesalahan memprediksi jumlah *potential customer*.

No.	Modeling	Train Precision	Test Precision
1.	KNN	0.82	0.52
2.	Decision Tree	1.00	0.72
3.	Random Forest	1.00	0.86
4.	AdaBoost	0.70	0.49
5.	XGBoost	1.00	0.80



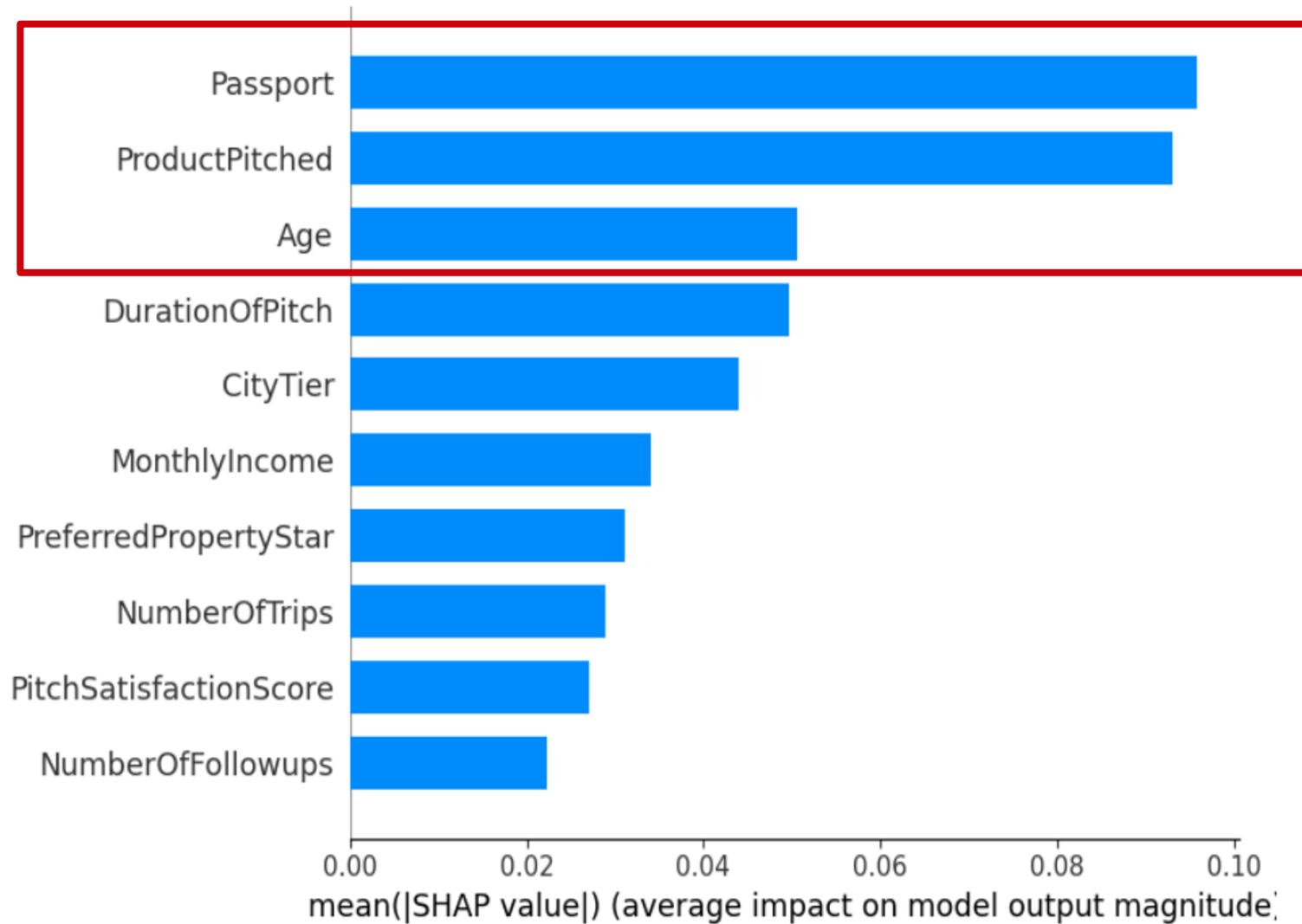
Hyperparameter Tuning

No.	Modeling	Train Precision	Test Precision
1.	Random Forest	0.9985	0.9405
2.	XGBoost	0.9994	0.9152

Random forest sebagai best-fit model karena selisih nilai data test dan data train sangat kecil dan tidak lebih dari **0,1**.



Feature Importance

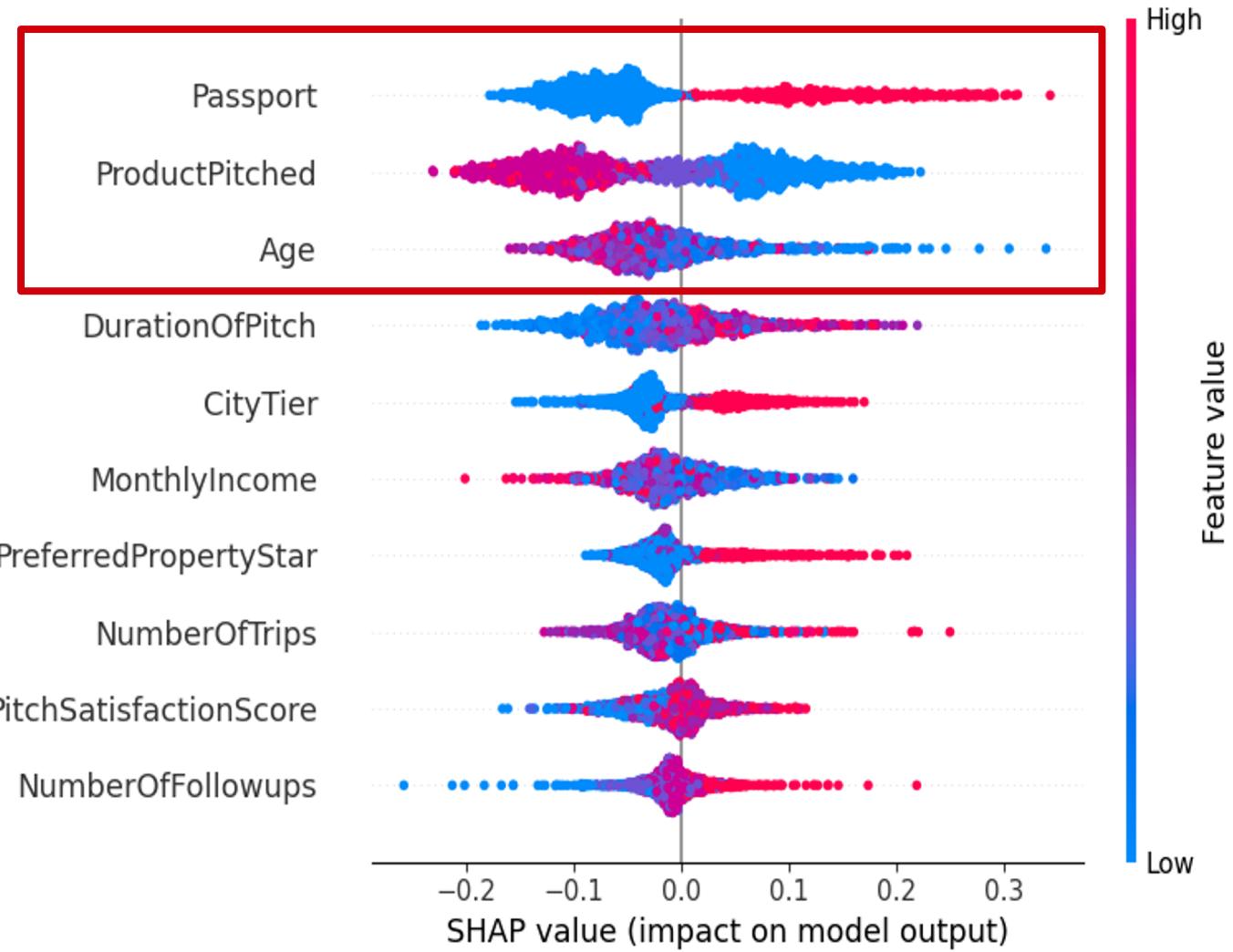


Dari hasil modeling random forest yang sudah dilakukan hyperparameter tuning, berikut fitur teratas yang paling penting yaitu **Passport**, **ProductPitched**, dan **Age**.



Features-Target Correlation

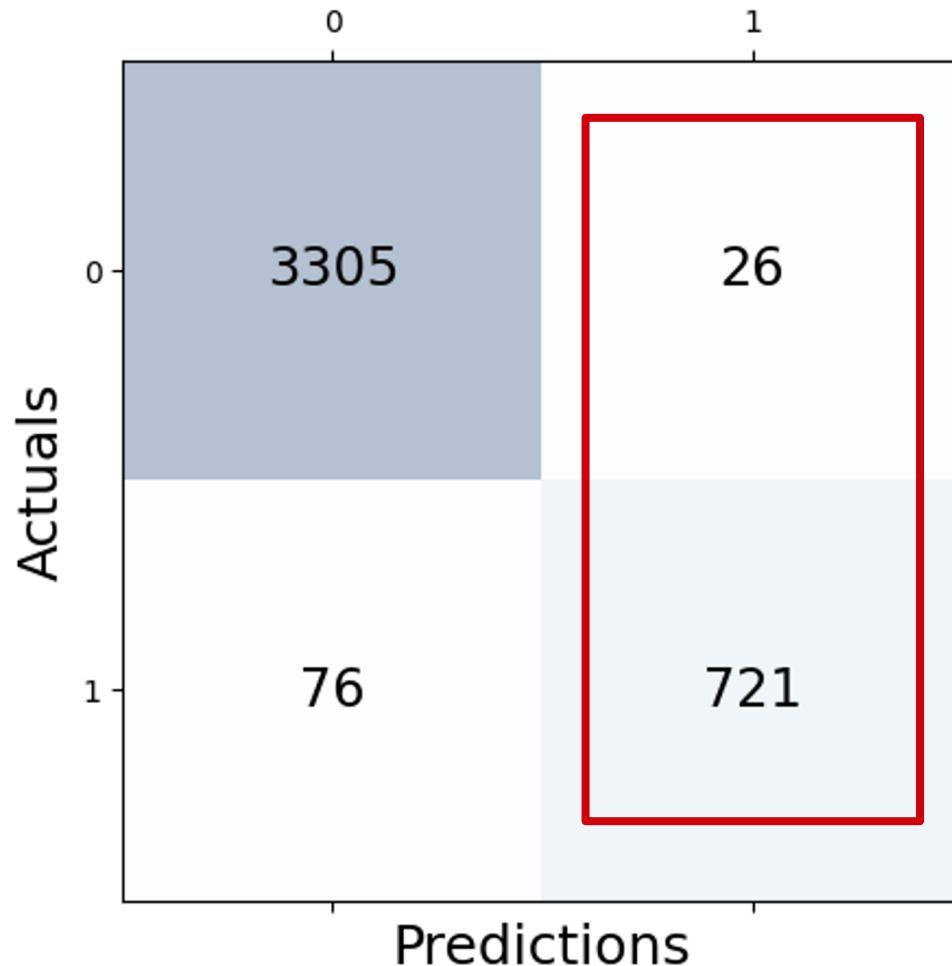
- *Passport* : Korelasi **positif**
- *ProductPitched* : Korelasi **negatif**
- *Age* : Korelasi **negatif**



BUSINESS SIMULATION



Confusion Matrix



- True Negative**
Diprediksi tidak mengambil paket liburan dan itu benar
3305
- False Negative**
Diprediksi tidak mengambil paket liburan dan itu salah
76
- False Positive**
Diprediksi mengambil paket liburan dan itu salah
26
- True Positive**
Diprediksi mengambil paket liburan dan itu benar
721

Kenaikan Total Revenue

Rp 215,1 Juta

Total Revenue (Before Modeling)

Rp 431,2 Juta

True Positive (TP)	: 920
False Positive (FP)	: 3968
Profit per package	: Rp
1.000.000,-	
Cost per marketing	: Rp
100.000,-	

Total Revenue (After
Modeling)

Rp 646,3 Juta

True Positive (TP)	: 721
False Positive (FP)	: 26
Profit per package	: Rp
1.000.000,-	
Cost per marketing	: Rp
100.000,-	

$$\text{Revenue} = (\text{TP} * \text{Profit per package}) - ((\text{TP} + \text{FP}) * \text{Cost per marketing})$$



Kenaikan Conversion Rate

78%

Conversion Rate (Before
Modelling)

19%

True Positive (TP) : 920
False Positive (FP) : 3968

Conversion Rate (After
Modelling)

97%

True Positive (TP) : 721
False Positive (FP) : 26

$$\text{Conversion Rate} = (\text{TP} / (\text{TP} + \text{FP})) \times 100\%$$





BUSINESS RECOMMENDATIO N



Penargetan Customer



Mendorong pelanggan untuk **memperoleh paspor** dengan melakukan penawaran khusus.

Misalnya dengan biaya yang sama dengan paket liburan dalam negeri, pelanggan yang memiliki paspor diberikan opsi destinasi wisata ke luar negeri dengan durasi liburan yang lebih lama atau akomodasi yang lebih bagus.

Penentuan Harga Paket Terbaru

Perusahaan dapat menentukan harga paket terbaru **menyesuaikan range harga paket kelas bawah hingga menengah** (Basic, Standard, dan Deluxe), sehingga dapat menarik perhatian pelanggan.





Paket Liburan Tematik

Menawarkan paket liburan tematik yang dirancang **khusus untuk anak muda** dapat menjadi daya tarik.

Misalnya, paket liburan petualangan ekstrem, tur kuliner kreatif, atau perjalanan musik dan festival. Sesuaikan paket liburan dengan minat dan gaya hidup generasi muda.





Pandas
LOVERS
Thank
You