

Final Presentation Batch 18

Holiday Package Prediction

PYLAB - Data Scientist (Group 7)



Us & our company

Trips&Travel.com

Sebuah perusahaan travel yang ingin menawarkan paket liburan baru: Wellness.

5 Paket liburan yang sudah ada:

- Basic
- Standard
- Deluxe
- Super Deluxe
- King



PYLAB

Sebuah tim data scientist dari divisi marketing Trips&Travel.com



PYLAB **Team**

**Reza
Kalmas**

Captain

**Feri Dwi
Saputro**

Member

**Fairuz Nur
Indah**

Member

**Cesilia
Natasya N**

Member

**Abram
Setyo P**

Member

Mardi

Member

**Yandri
S R**

Member

Table **of contents**

01

**Business
Understanding**

02

**Data Overview,
EDA & Insight**

03

**Data Cleaning &
Preparation**

04

**ML Modelling &
Evaluation**

05

**Business Insight &
Recommendation**



01

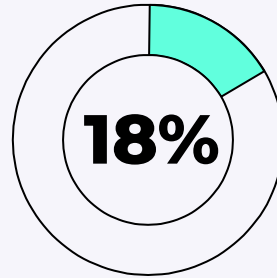
Business Understanding

Problem, goal, objectives, and
business metrics





Problem Statements



Pelanggan

membeli paket liburan, namun masih dihubungi secara acak.

Perusahaan ingin memanfaatkan data yang tersedia untuk meluncurkan paket yang baru.

Goals

- ★ Memprediksi pelanggan yang akan membeli paket liburan baru untuk efisiensi pengeluaran perusahaan.
- ★ Menganalisa karakteristik pelanggan potensial menggunakan data tahun lalu

Objectives

- ★ Membuat sistem untuk memprediksi pelanggan yang akan ditargetkan saat marketing campaign agar pengeluaran biaya lebih efisien.
- ★ Mengetahui karakteristik pelanggan potensial yang akan membeli paket liburan.

Business Metric

- ★ Conversion rate, Marketing Cost

02

Data Overview

EDA & Insight



Data Overview

Profile Customer:

CustomerID
Age
CityTier
Occupation
Gender
NumberOfPersonVisiting
PreferredPropertyStar
MaritalStatus
NumberOfTrips
Passport
OwnCar
NumberOfChildrenVisiting
Designation
MonthlyIncome

Information of Pitch:

TypeofContact
NumberOfFollowups
DurationOfPitch
ProductPitched
PitchSatisfactionScore

Target:

ProdTaken

- ❖ **4888 baris**
- ❖ **20 kolom**
- ❖ 14 kolom numerik
- ❖ 6 kolom kategorik

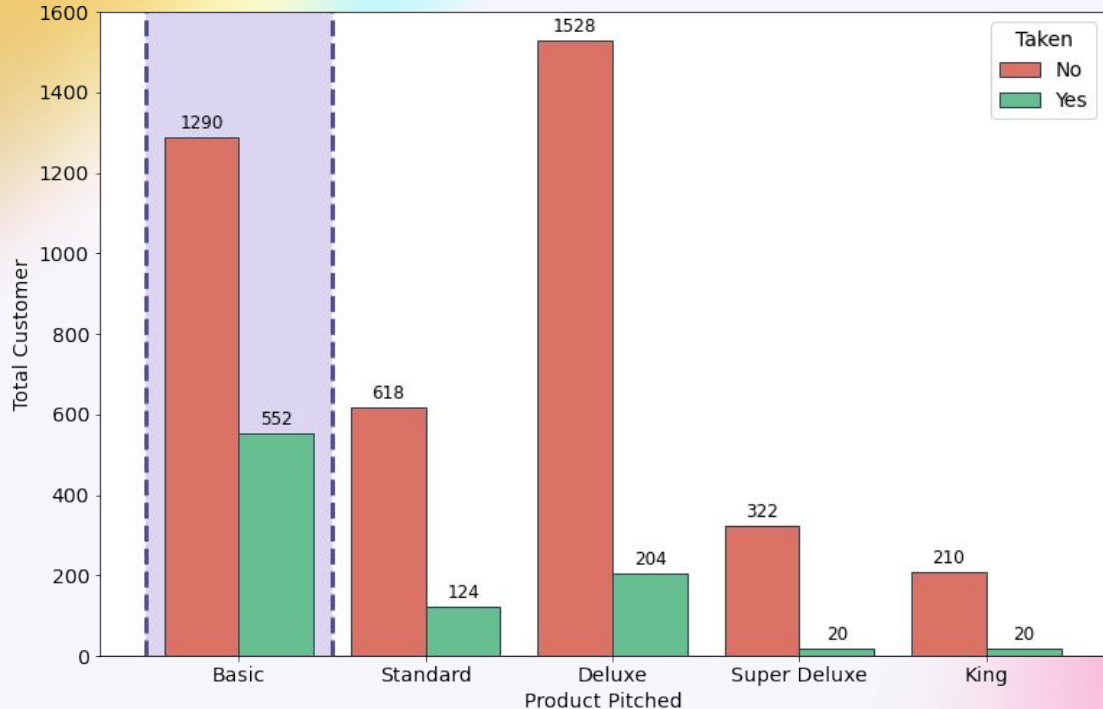


EDA ⁺ Insights

Exploratory Data Analysis

Customers are More Interested in Basic Products than Other Products

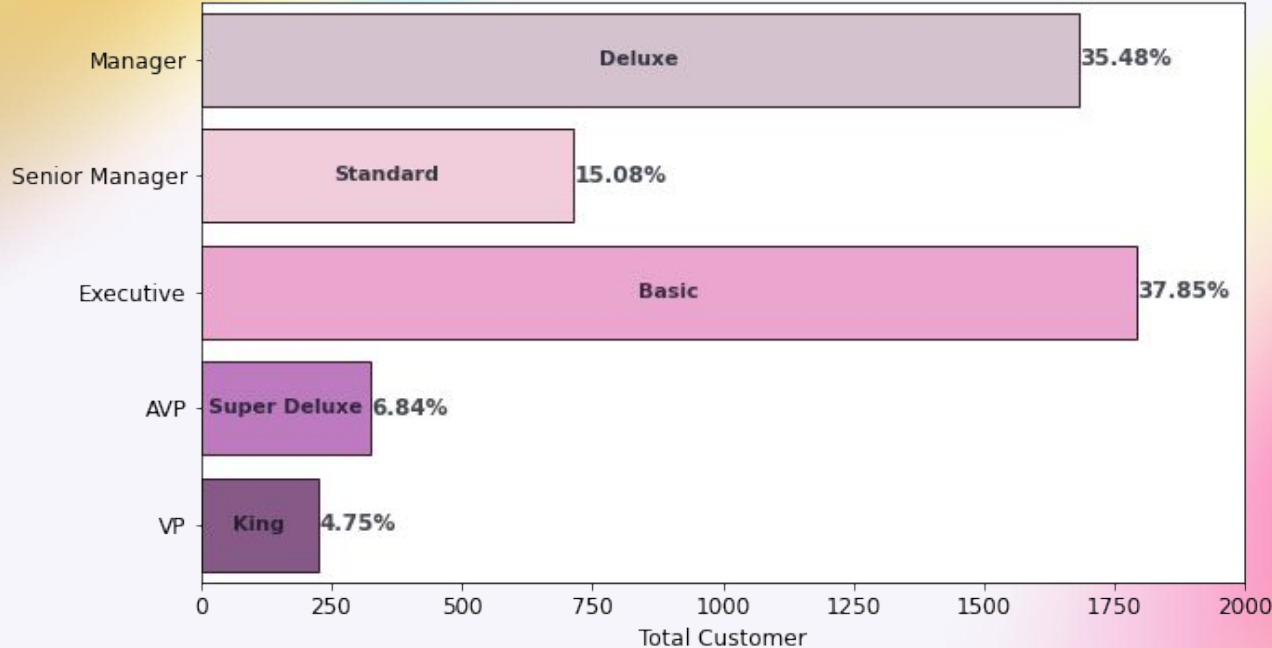
We can offer promos or discounts on products other than basic during product pitching so that customers are also interested in buying them



Customer **lebih tertarik membeli paket Basic** dibandingkan paket lainnya, maka kita perlu **memberikan promo atau diskon terhadap paket lainnya** agar customer menjadi tertarik juga.

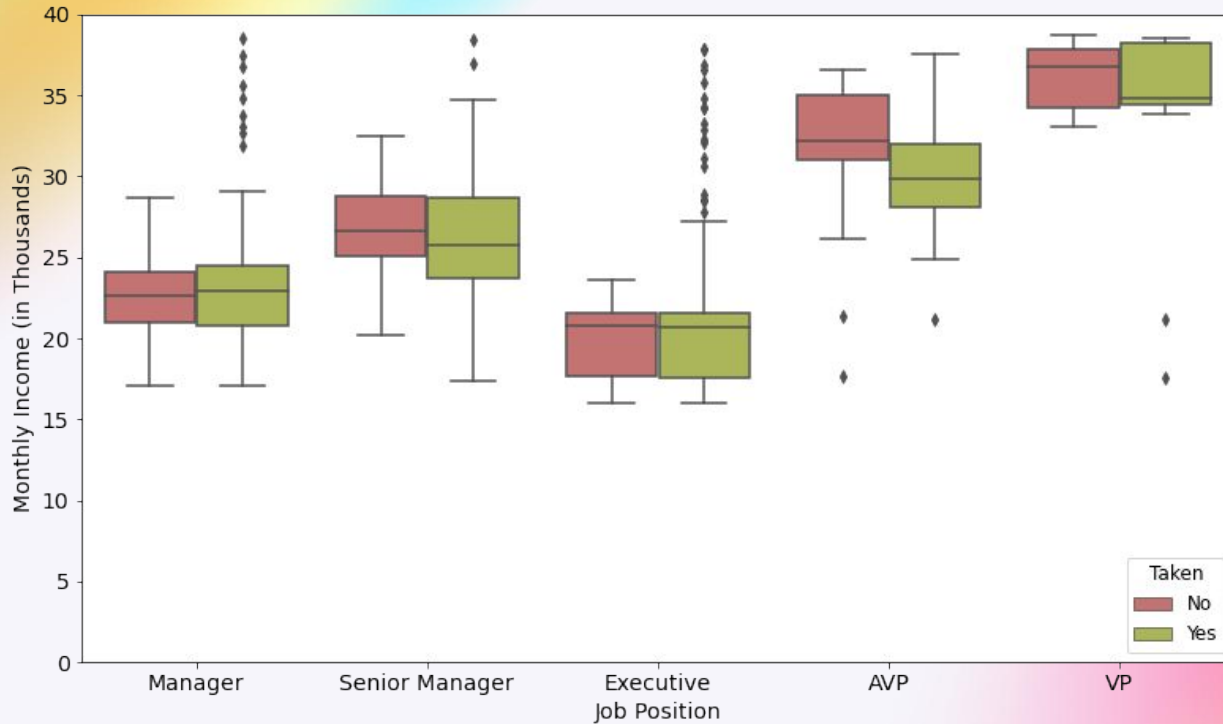
Customers are Only Offered One Type of Product in Each Job Title

We Can Offer Other Types of Products as well during Pitching
which Allows Customers to be More Interested in Buying These Types of Products



Tiap Jabatan hanya ditawarkan 1 jenis product, seharusnya pelanggan dapat diberikan pilihan yang lebih banyak yang kemungkinan tertarik juga.

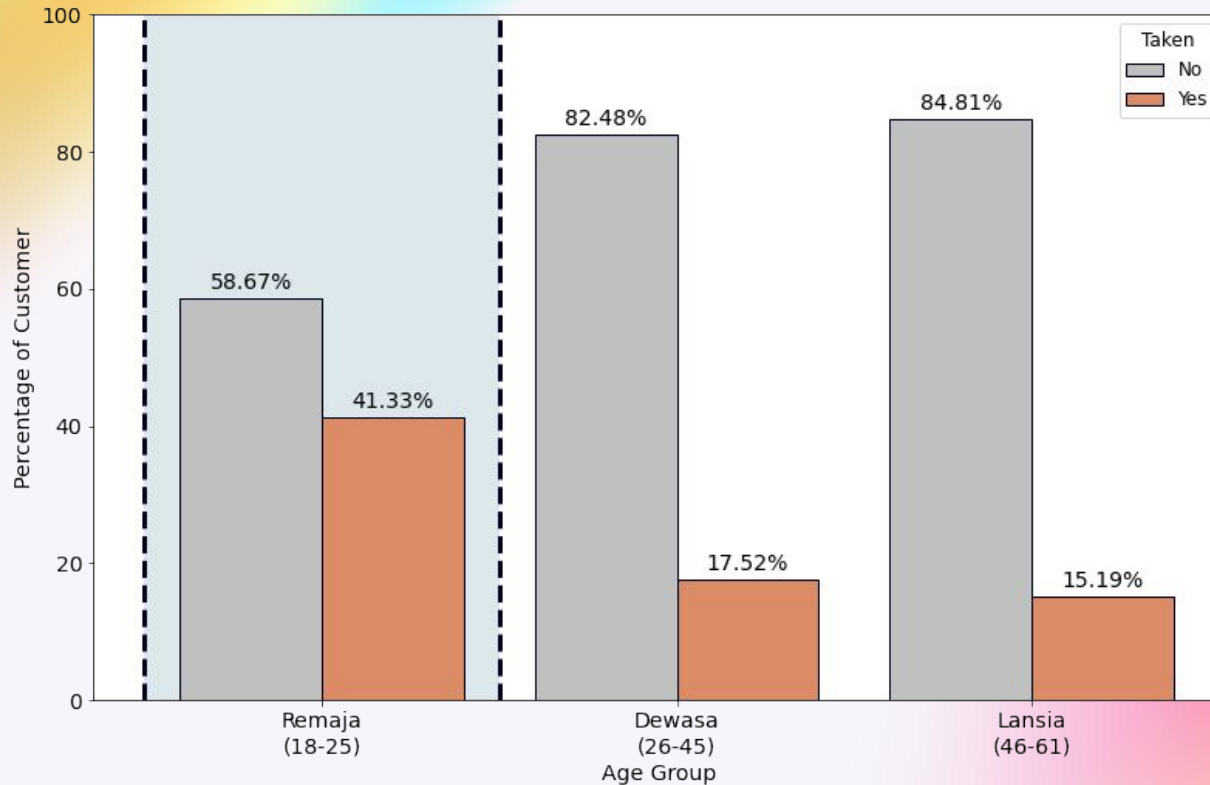
Customers with Higher Monthly Income Will be Offered Better Products too



Ternyata jenis produk yang ditawarkan hanya berdasarkan montly income, **semakin tinggi maka jenis produk yang ditawarkan lebih bagus juga.**

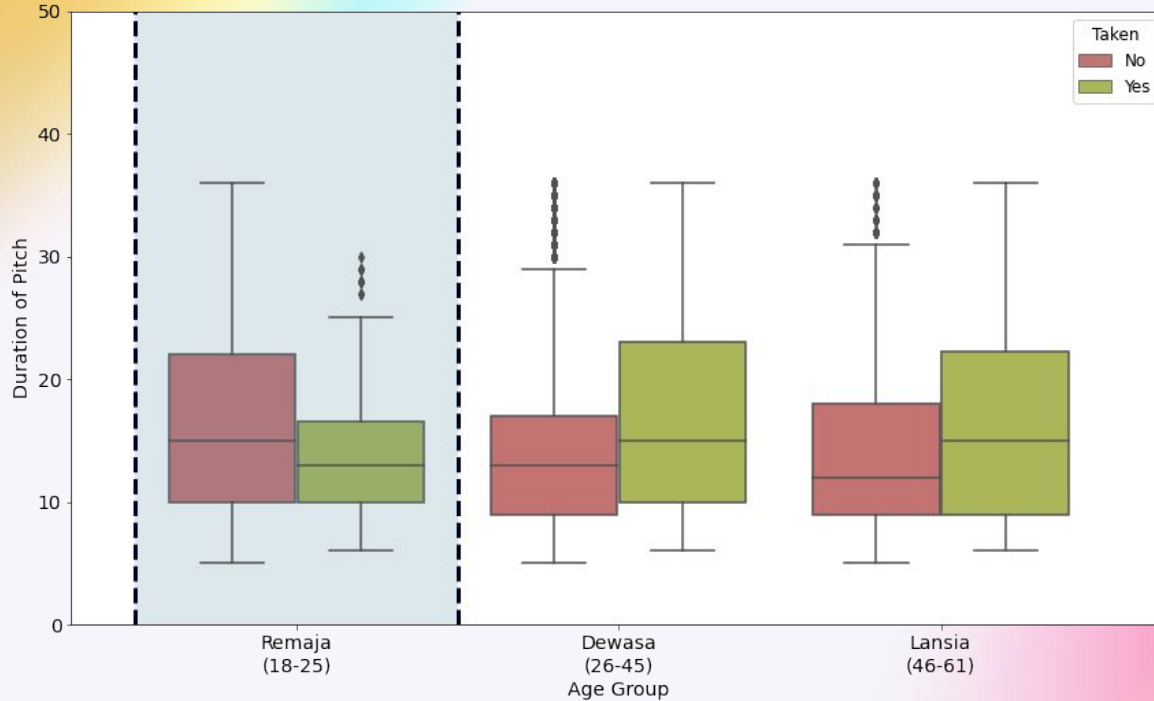
Around 40% of Teenage Customers are More Likely to Purchase Holiday Packages

Customers who are 17 to 25 Years Old Can be Prioritized First During The Marketing Campaign



Kelompok umur Remaja (18-25 tahun) cenderung lebih tertarik membeli paket liburan sehingga bisa diprioritaskan terlebih dahulu.

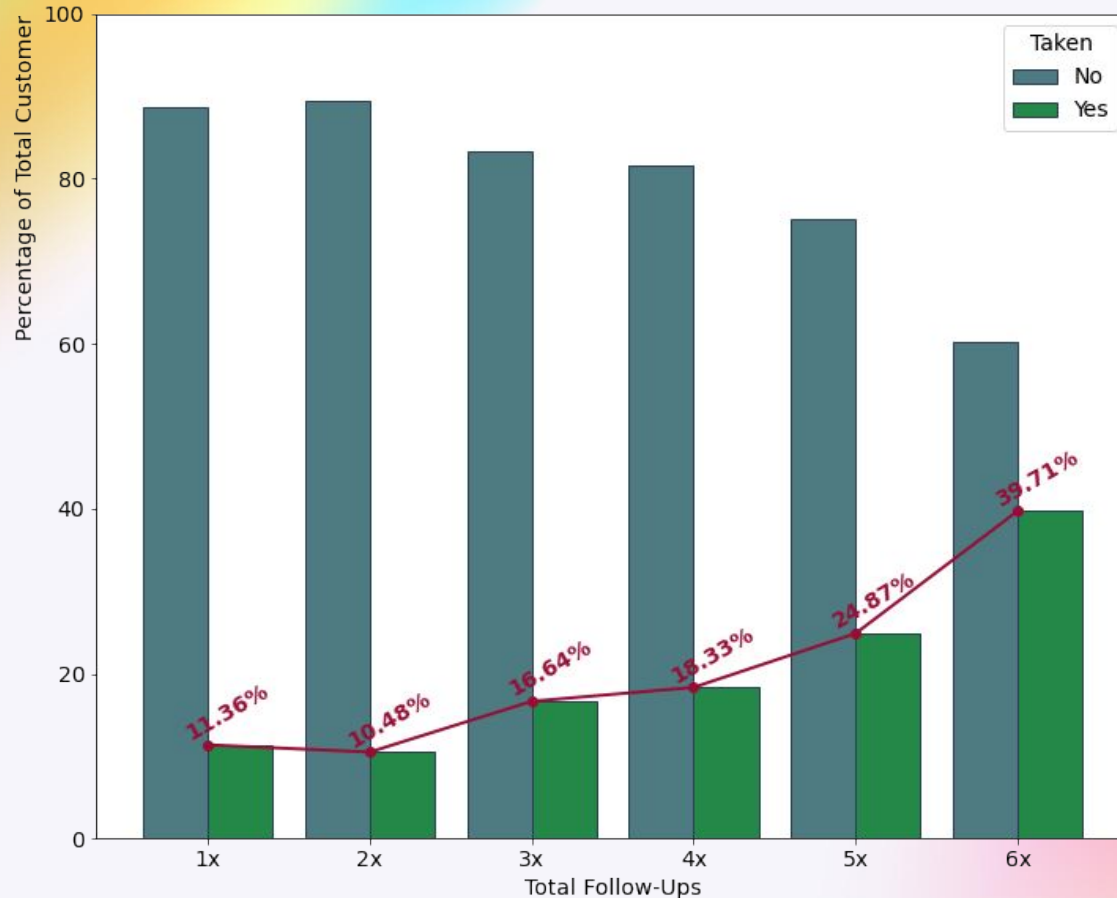
The Duration of the Pitch for Customers who Buy Holiday Packages Based on Age Group and Product Type



Pelanggan dengan kelompok umur adalah **Remaja (18-25 tahun)** cenderung membeli paket dengan **durasi di bawah 20 menit**.

Follow Up After Pitching the Product at Least 3 Times or More

With More Follow-Ups, Customers Have a Higher Chance to Buy Travel Packages



Semakin banyak follow up yang dilakukan kepada customer , maka **semakin besar kemungkinan customer membeli** paket liburan.

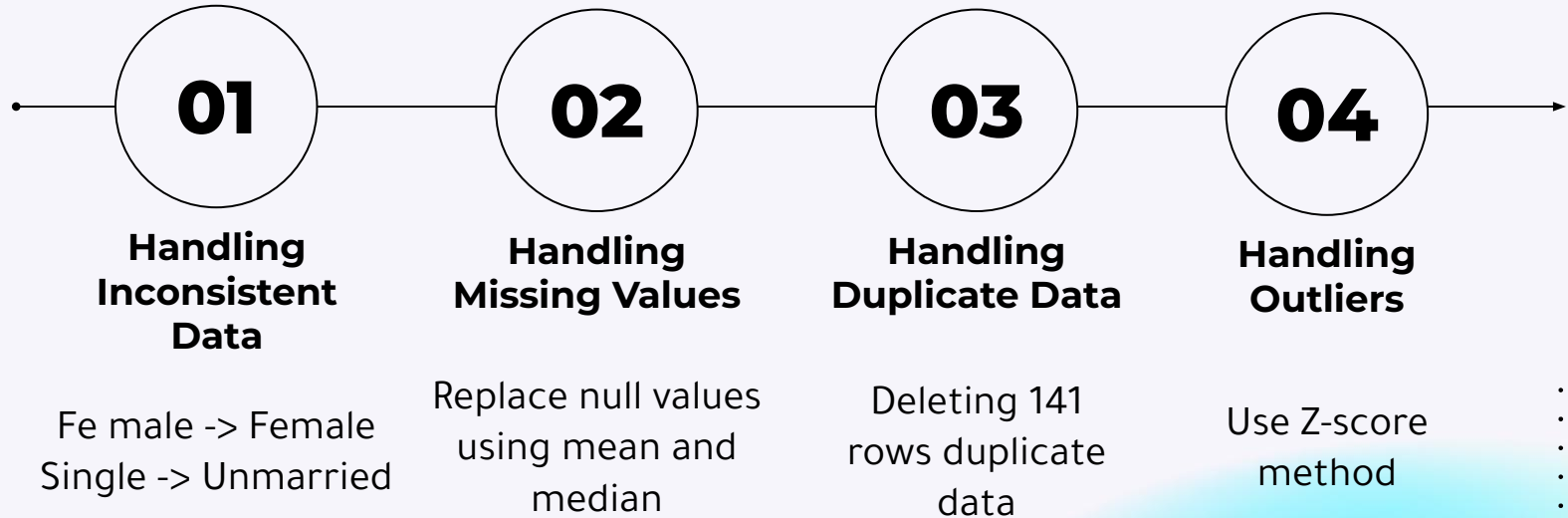
03

Data Cleaning & Preparation

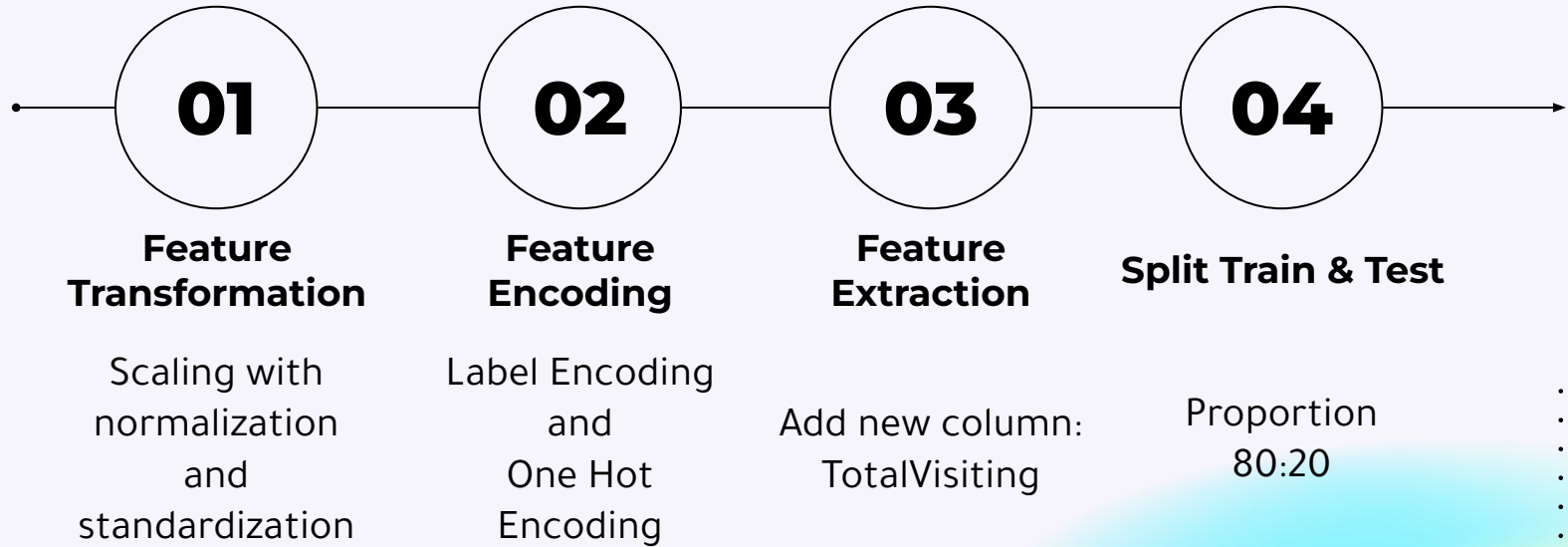
Clean and prepare the data +
feature engineering



Data Cleaning



Feature **Engineering**

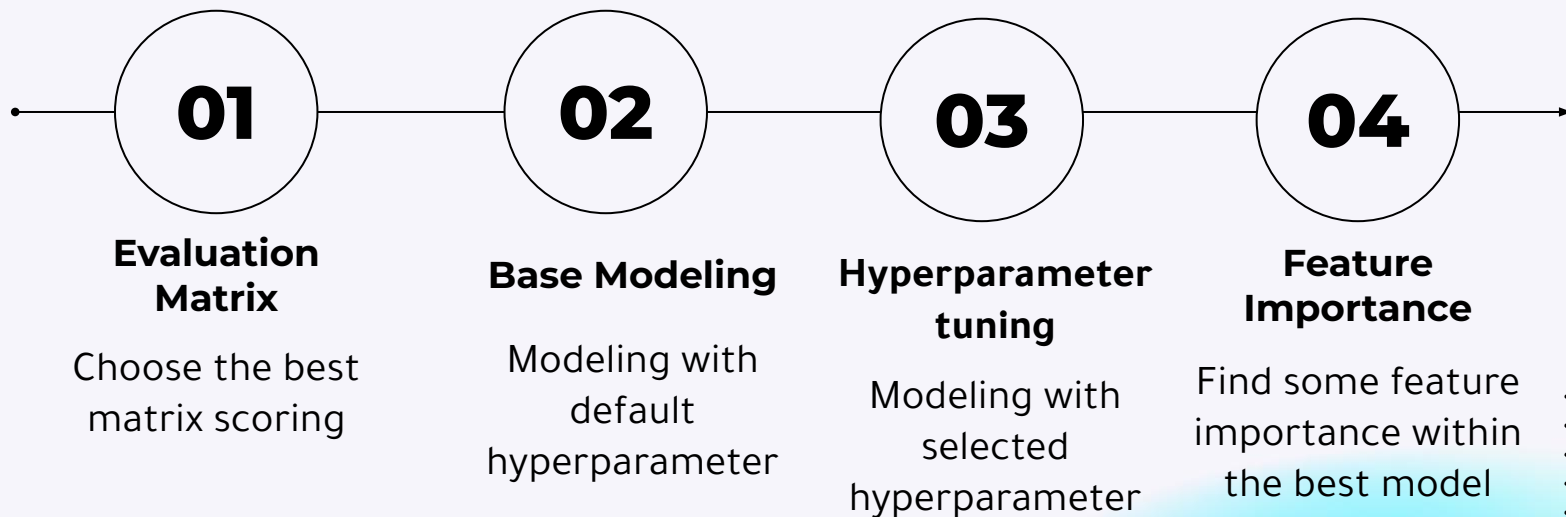


04

ML Modelling & Evaluation



ML modeling & **Evaluation**



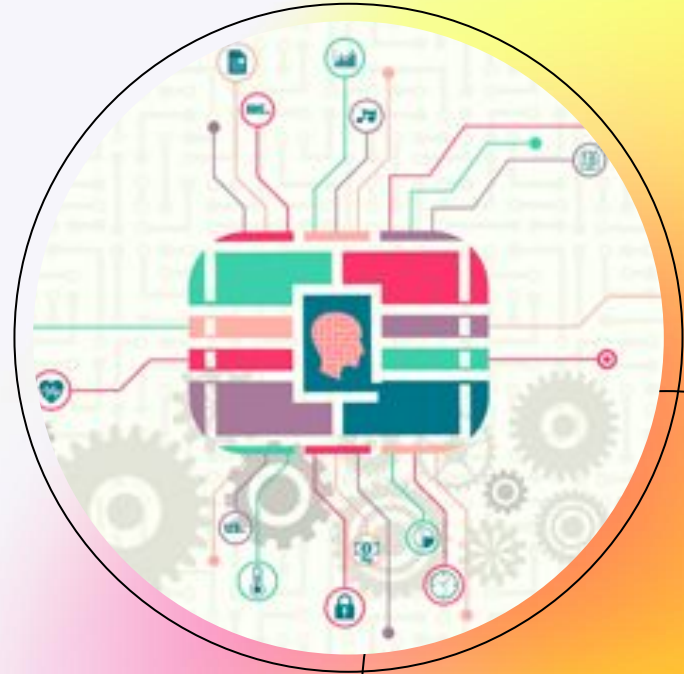
Evaluasi **Matrix**

Primary matrix: F1 Score

Untuk mempertimbangkan hasil dari Precision dan Recall

Secondary matrix: AUC

Untuk pendukung untuk mengukur apakah model bisa membedakan pelanggan antar kelas.



Data Modelling

Menggunakan 19 fitur dari feature yang ada

Model After Hyperparameter Tuning

	Model	Accuracy	Precision	Recall	F1 Score	MCC	AUC (Test)	AUC (Train)
0	Logistic Regression	0.8564	0.7361	0.3118	0.4380	0.4160	0.8362	0.8041
1	Decision Tree	0.8933	0.6845	0.7529	0.7171	0.6527	0.8385	1.0000
2	Random Forest	0.9398	0.8897	0.7588	0.8190	0.7868	0.9754	1.0000
3	AdaBoost	0.8501	0.6944	0.2941	0.4132	0.3849	0.8509	0.8364
4	XGBoost	0.9409	0.8851	0.7706	0.8239	0.7913	0.9707	1.0000

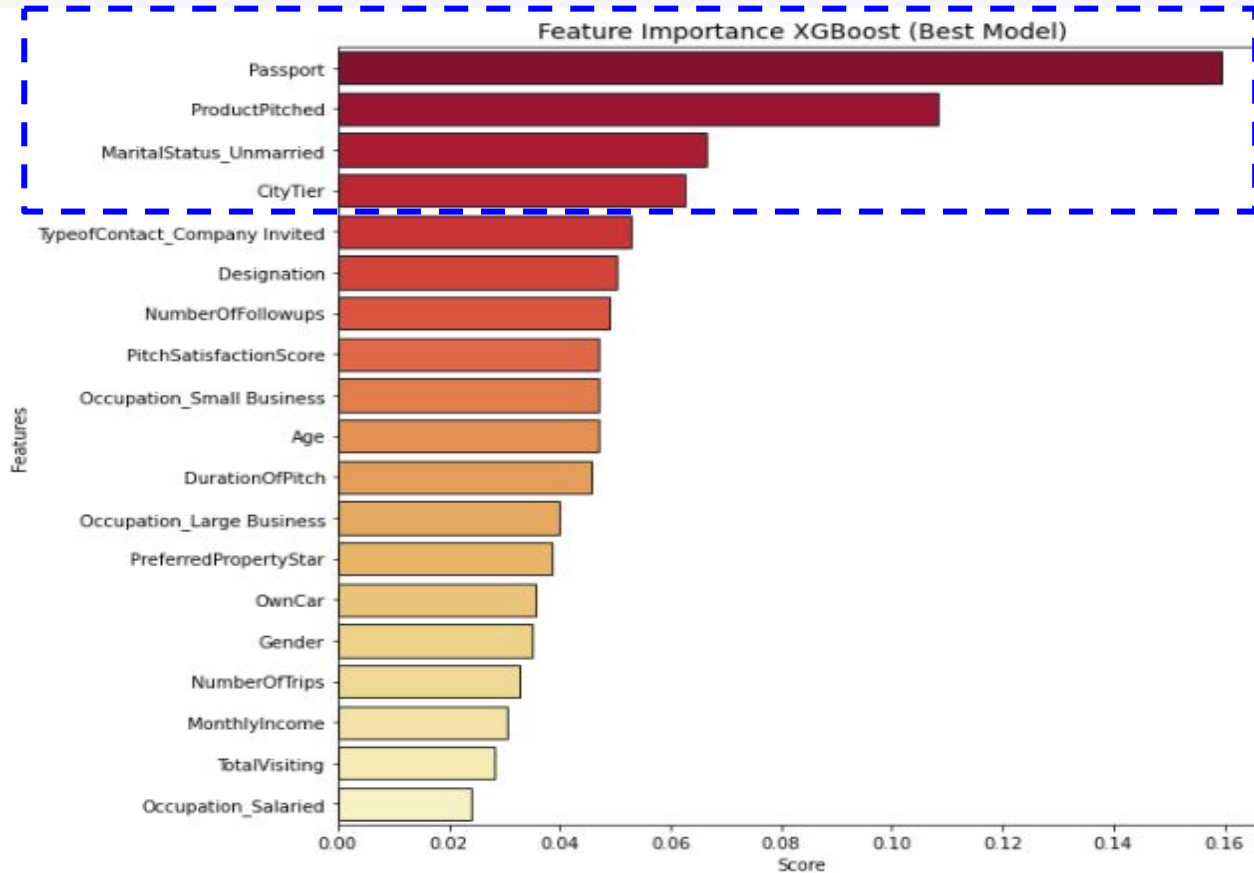
XGBoost menjadi model yang terbaik dengan F1 Score paling tinggi yaitu 82.39% dengan score confusion matrix nya **TP 131, FP 17, TN 760, FN 39** dan mempertimbangkan hasil scoring metric yang lain setelah hyperparameter tuning yang dilakukan.

05

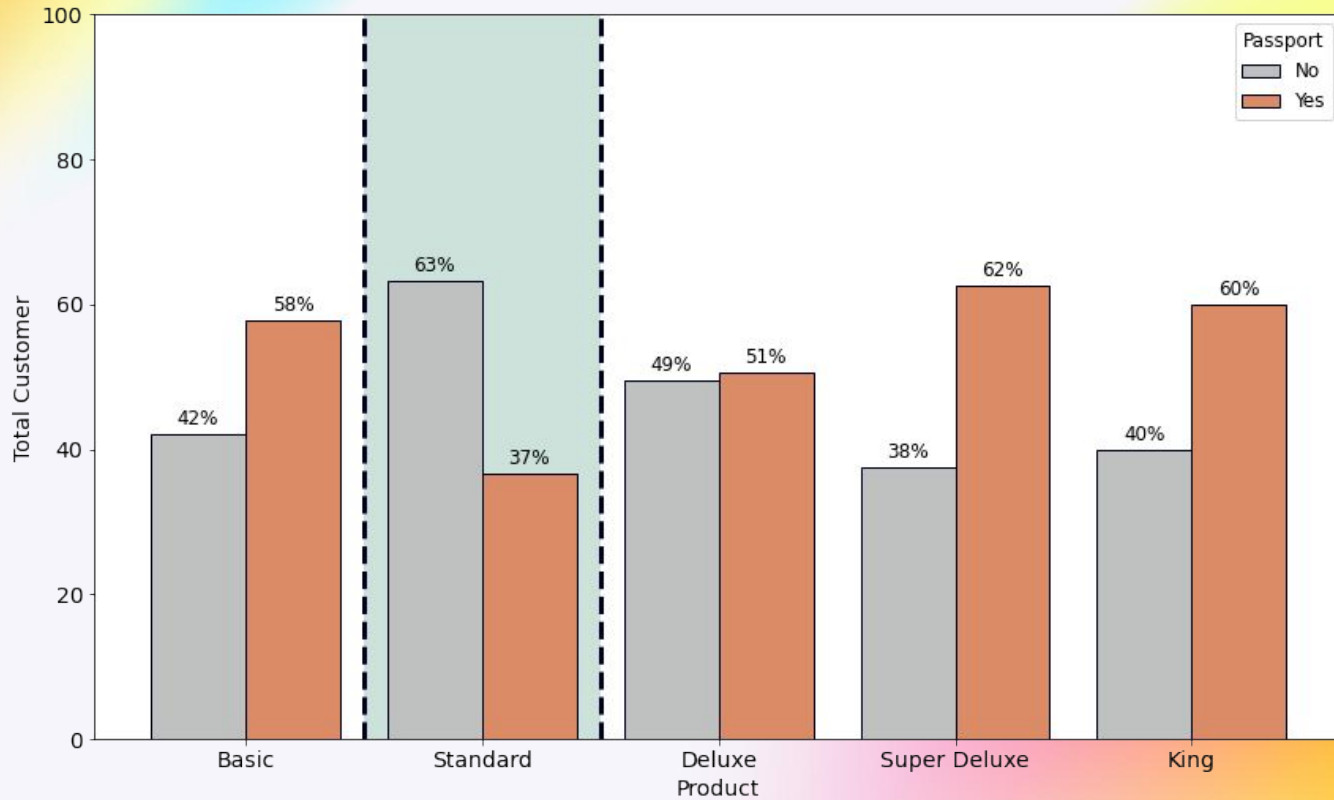
Business Insight & Recommendation



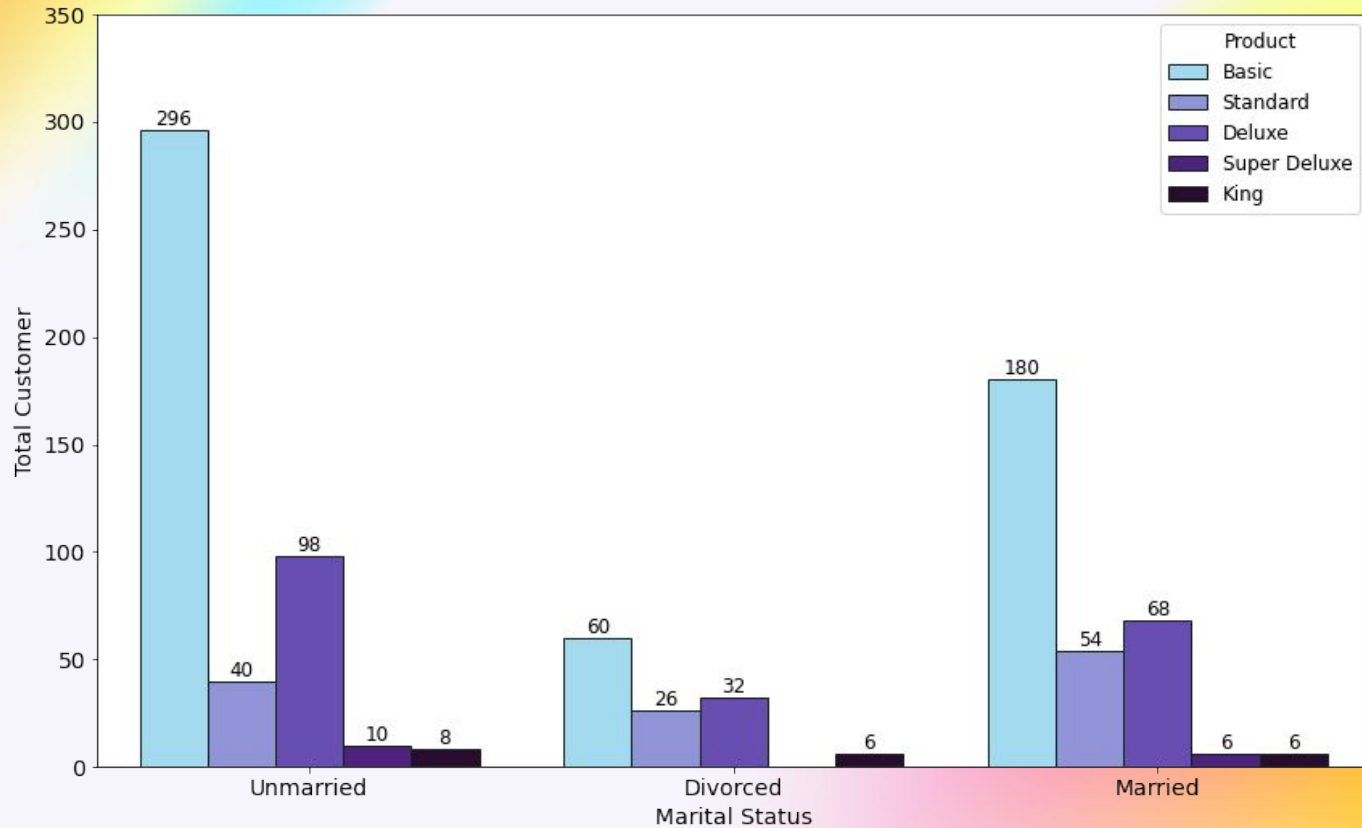
Feature Importance with XGBoost



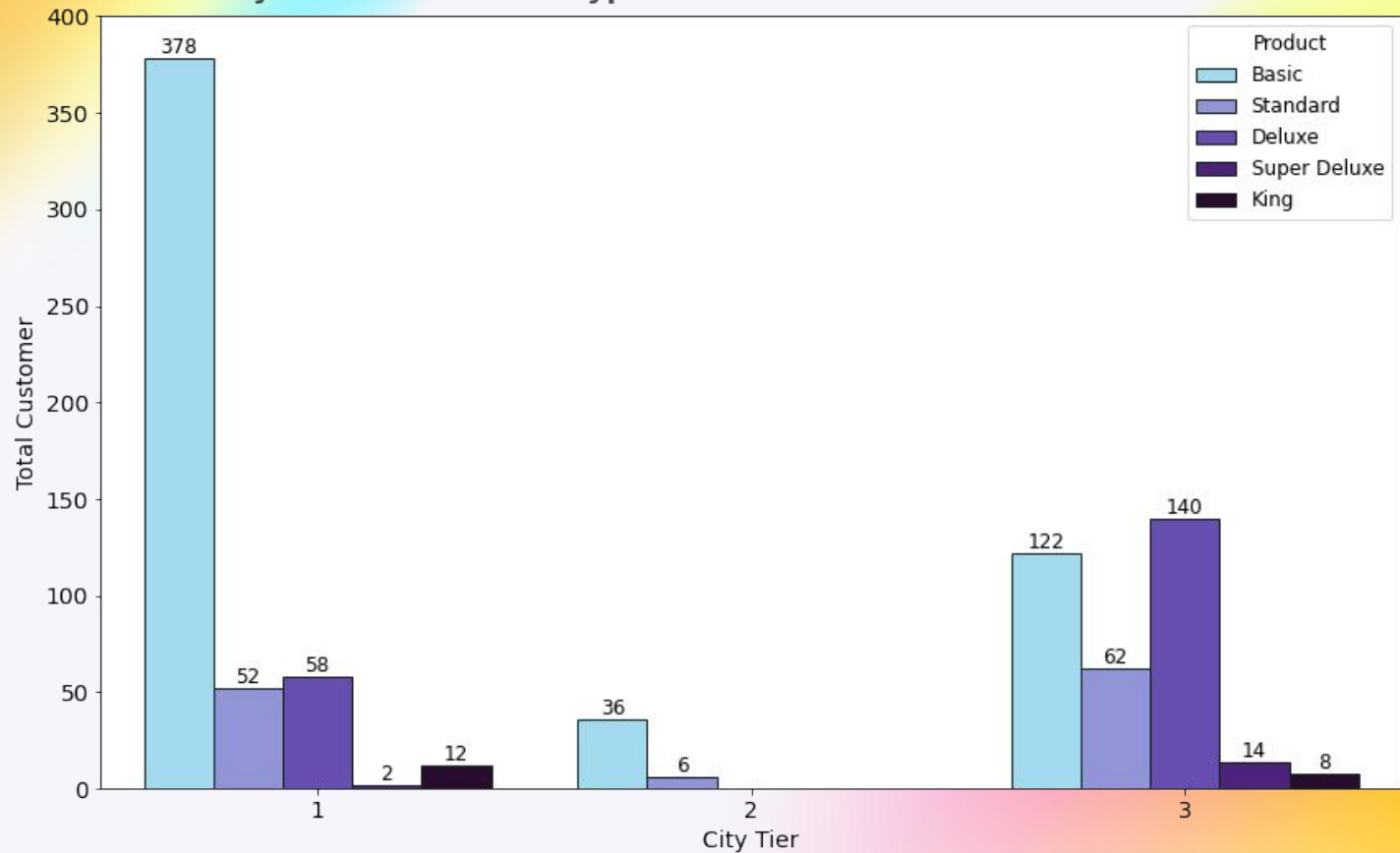
The Number of Customers who Buy Holiday Packages Based on Product Type and Passport Ownership



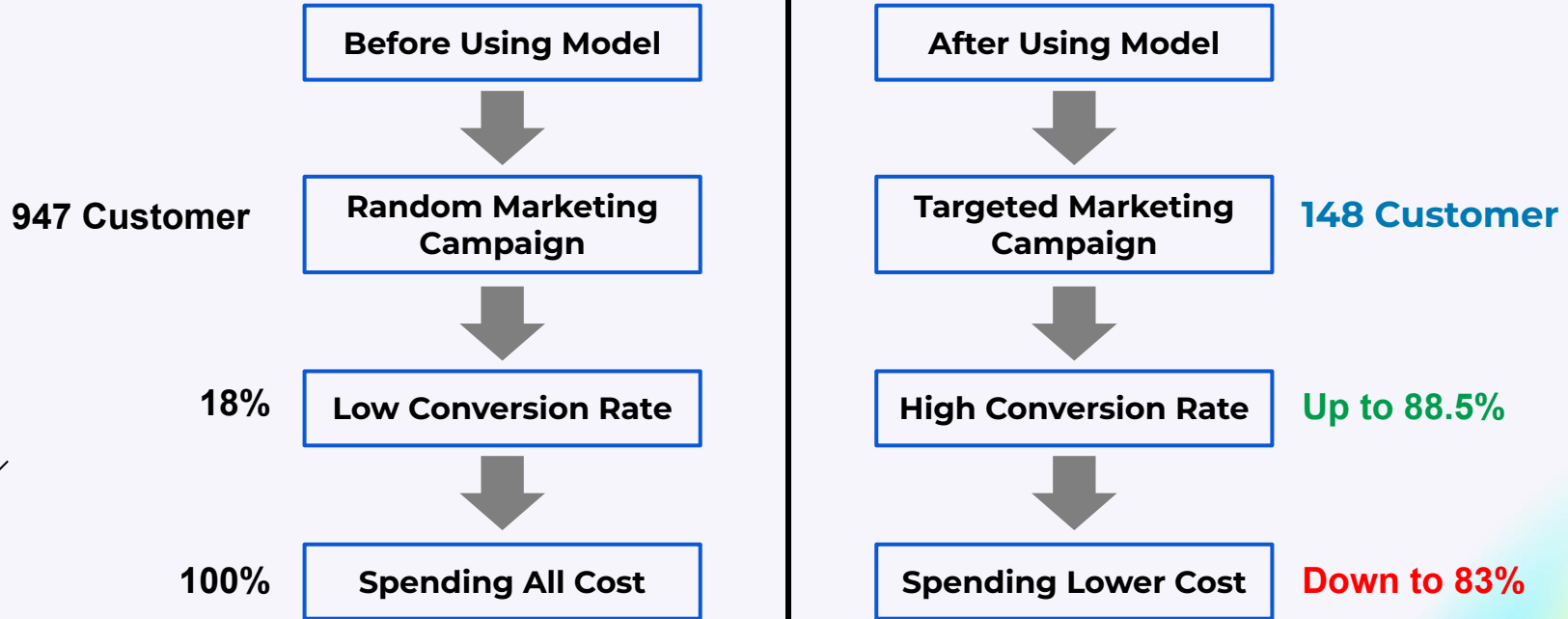
**The Number of Customers who Buy Holiday Packages
Based on Marital Status and Product Type**



**The Number of Customers who Buy Holiday Packages
Based on City Tier and Product Type**





Business Workflow (With Test Data)



Jika diasumsikan



Business Impact (with Test Data)

	Before Using Model	After Using Model
Total Customer	947	148
Targeted Customer	170	131
Conversion Rate	18%	 88.5%
Marketing Cost	\$42342 <small>*Total Duration of Pitch (all class) x cost per minute</small>	\$42342 <small>*Total Duration of Pitch (all class) x cost per minute</small>
Spending Cost	\$42342	\$7245
Saving Cost	0%	 83%
Revenue	\$ 30300 <small>*Total (Number of Product Purchased x Price x 10%)</small>	\$ 23500 <small>*Total (Number of Product Purchased x Price x 10%)</small>

Revenue from All Budget

	Before Using Model	After Using Model	%	After Model and All Budget	%
Revenue	\$ 30300	\$ 23500	23	\$75000* (23500+51500)	247.5



Asumsi:

Menggunakan seluruh sisa budget : \$35097 (all cost - spending cost)
Cost per orang (\$3 x 20 menit) : \$60
Dapat melakukan telemarketing : 585 orang
Dengan CvR 88%, asumsi membeli : 515 orang

Jika yang terjual produk basic maka akan meningkatkan revenue **\$51500**

Business Recommendation

Melakukan **marketing campaign** terhadap **potential customer** untuk **paket Wellness** dengan ketentuan:

1. Jika **paket Wellness** terdiri dari beberapa jenis maka bisa **diprioritaskan dari yang paling murah sesuai dengan jabatan** pekerjaan pelanggannya.
2. Prioritaskan dahulu pada kelompok umur **Remaja (18-25 tahun)** dengan **durasi pitch di bawah 20 menit**.
3. **Paket Wellness** paling murah bisa ditawarkan pada pelanggan dengan status **unmarried** atau tinggal di **city tier 1 dan 2**.
4. **Paket Wellness** yang medium bisa ditawarkan pada pelanggan yang **belum memiliki passport** atau tinggal di **city tier 3**.
5. **Paket Wellness** lainnya bisa ditawarkan sesuai dengan **montly income** pelanggan tersebut.
6. Melakukan **follow up lebih dari 3 kali**.

Business **Recommendation**



Jika menggunakan model, revenue yang didapatkan lebih sedikit, maka kami **merekomendasikan marketing cost yang tersisa** untuk digunakan **mencari basis pelanggan baru** sehingga revenue yang dihasilkan meningkat.

Thank you!

