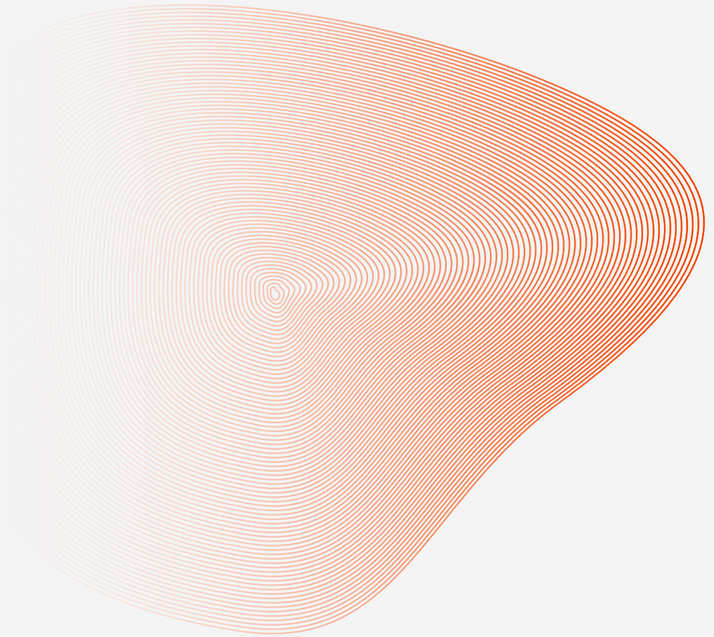
Data Traveller

Potential Customer **Prediction**

Data Traveller

Agenda

- Team Introduction
- Industry Analysis
- Problem Statement
- Our Solutions



Data Traveller

Our **Team**

BACK TO AGENDA PAGE

Data Team of Chikitrans

We are internal data team in a company called Chikitrans, travel agency who sells travel package.

We alongside with marketing team, are responsible for providing business recommendations based on available data from marketing team to improve the sales performance of the company.

Tony Hermawan Widjanarko

Project Leader & Data Scientist

Ryan Anugrah

Business Analyst

Esraminar Siregar

Data Scientist

Farhan Rizki

Business Analyst

Rayhan Prawira Daksa

Data Scientist

Rianita

Data Analyst

Data Traveller

Industry **Analysis**

85% of respondents

surveyed say they plan to take two or more leisure trips in 2023

74% of respondents

surveyed agree that they care more about creating a travel experience that meets their expectations than about the cost

78% of respondents

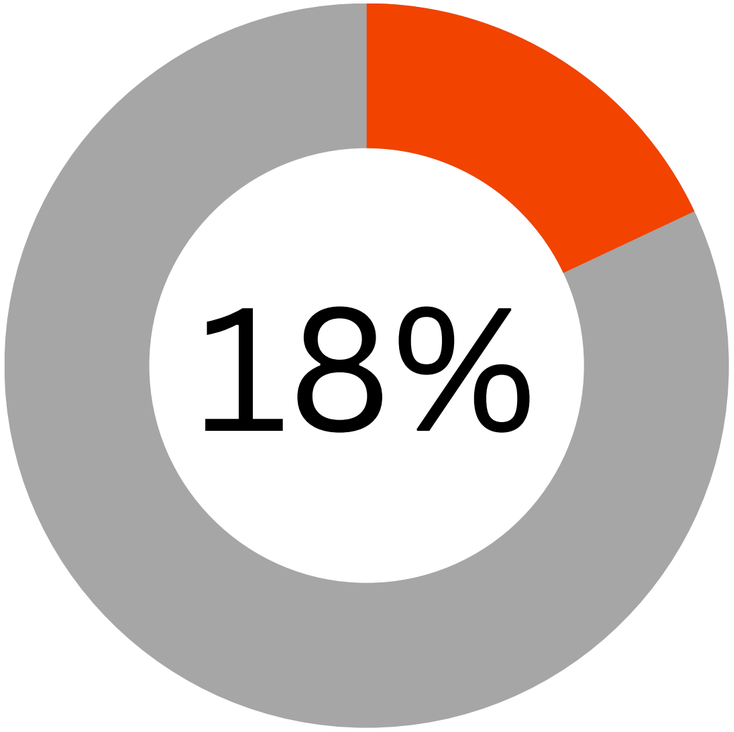
surveyed agree they see leisure travel as an important budget priority

BACK TO AGENDA PAGE

Source : American Express Travel - 2023 Global Travel Trends Report

Data Traveller

Problem **Statement**



Conversion Rate

Based on sales data that has been recorded so far, the **conversion rate** obtained is **only at 18%**.

Marketing costs incurred to follow up the prospective customers are also high, namely **IDR 10,000 per follow up**.

out of

4,888

customers pitched

BACK TO AGENDA PAGE

BACK TO AGENDA PAGE

Goals	Objective	Business Metric
Increase conversion rate from primary business (travel package sales)	Create a classification model to predict potential customers to increase conversion rate	Conversion Rate Customer Acquisition Cost

Our **Solutions**

Data Traveller

Preliminary **Insight**

Dataset

There are 4,888 rows of data that consist of customer profile and transaction details.

BACK TO AGENDA PAGE

Customer Profile

- CustomerID
- Age
- CityTier
- Occupation
- Gender
- MaritalStatus
- Passport
- OwnCar
- MonthlyIncome
- Designation

Transaction Details

- TypeofContact
- DurationOfPitch
- NumberOfPersonVisiting
- NumberOfFollowups
- ProductPitched
- PreferredPropertyStar
- NumberOfTrips
- PitchSatisfactionScore
- NumberOfChildrenVisiting

Target Column

- ProdTaken

Data Traveller

Preliminary **Insight**

Exploratory Data Analysis

Which occupation who did purchase

The majority who buy the package are **employees/entepreneurs**

BACK TO AGENDA PAGE

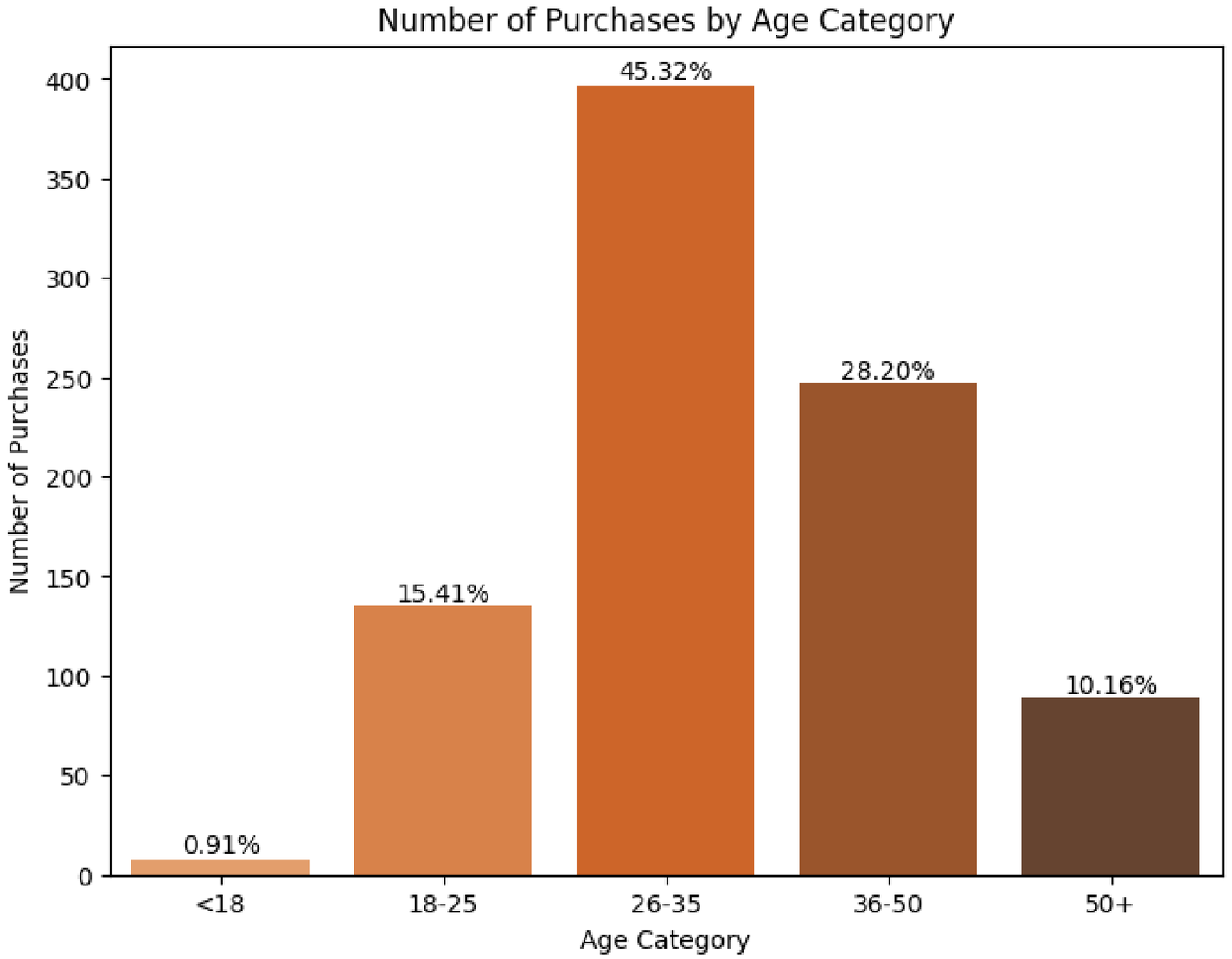
Data Traveller

Preliminary **Insight**

Exploratory Data Analysis

Purchase by Age Category

BACK TO AGENDA PAGE



The majority who buy the package are **productive individual** with **age ranged from 26 to 35** years old

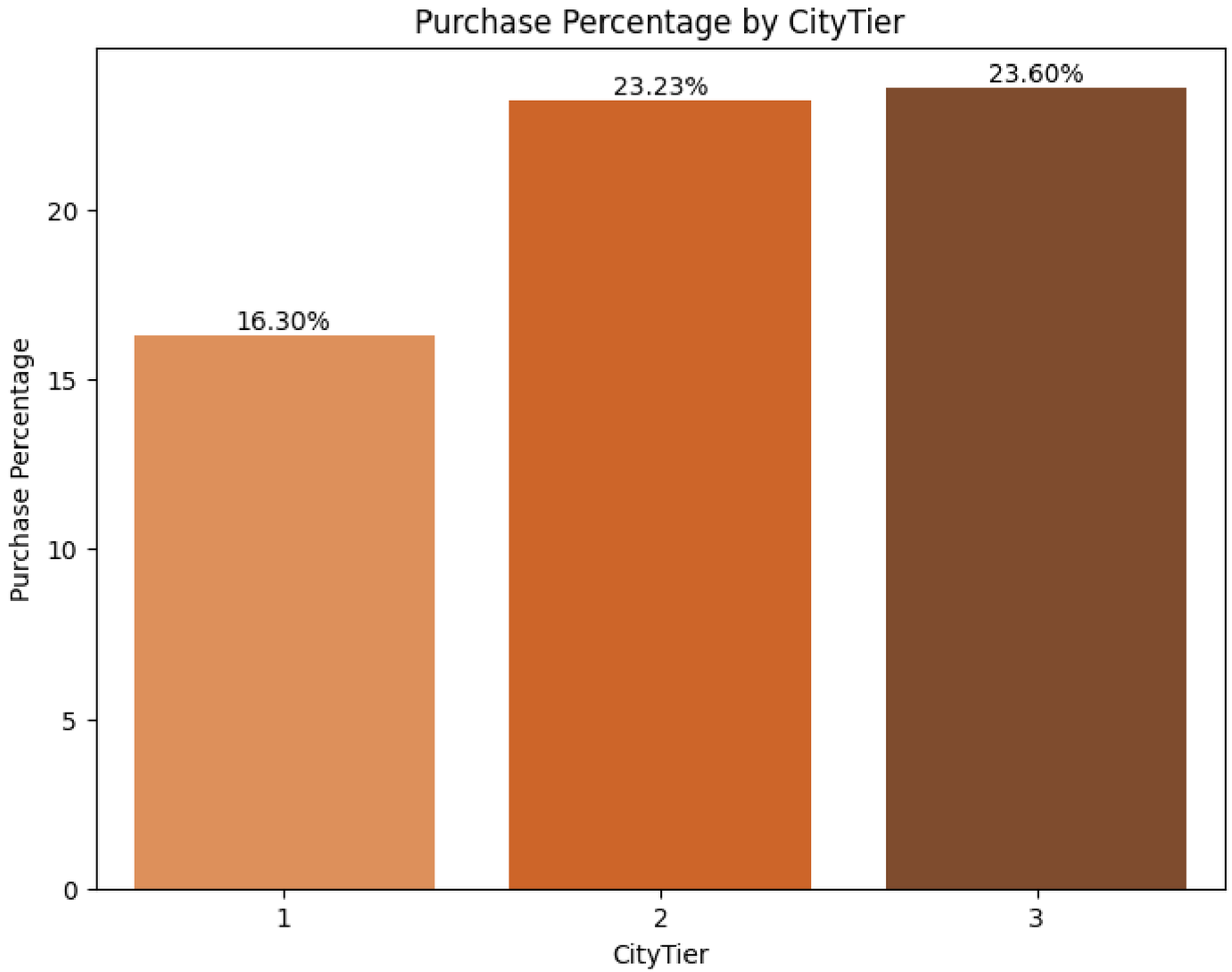
Data Traveller

Preliminary **Insight**

Exploratory Data Analysis

Purchase by City Tier

BACK TO AGENDA PAGE



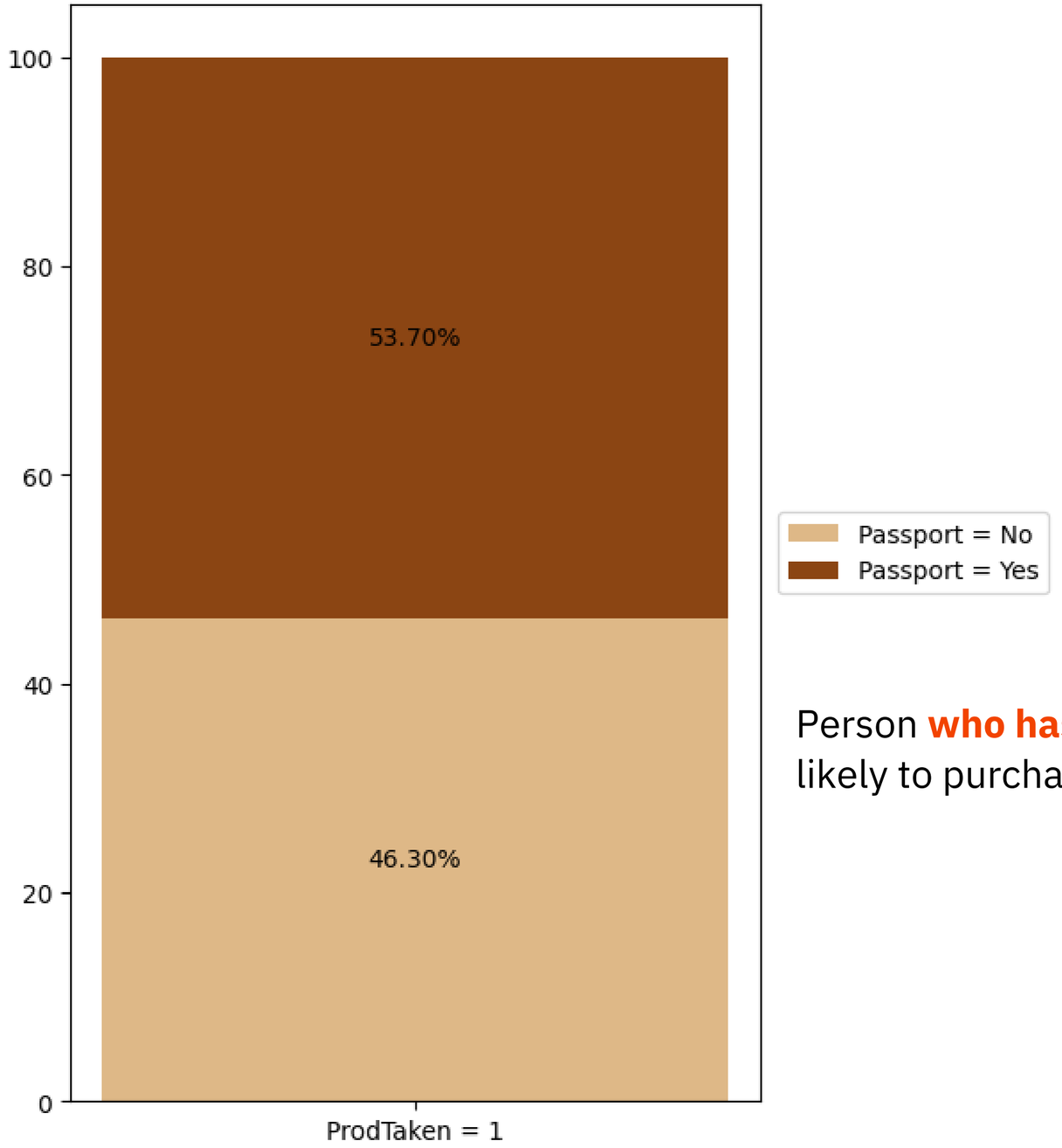
People reside in **Tier 2 & Tier 3** has more purchasing power than Tier 1 population

Data Traveller

Preliminary **Insight**

Exploratory Data Analysis

Purchase by Passport Holder



Person **who has passport** most likely to purchase

BACK TO AGENDA PAGE

BACK TO AGENDA PAGE

Data **Preprocessing**

Missing, Invalid Values & Duplicated Data	We handled missing values in columns Age, DurationOfPitch, NumberOfFollowups, PreferredPropertyStar, NumberOfTrips, NumberOfChildrenVisiting, MonthlyIncome. We also handle invalid values in columns Gender and MaritalStatus. Lastly, we dropped the duplicated data.
Outlier Handing, Feature Transformation & Feature Encoding	We handle outliers with z-score and transform the feature using standardization/logarithm to make sure all features are stable for modelling. Categorical columns are encoded using LabelEncoder or One Hot Encoding.
Feature Selection	After looking at ANOVA result, we decided to include all columns as feature because there is no big gap between one another in terms of importance. We dropped several columns in DurationOfPitch, NumberOfFollowups, ProductPitched and PitchSatisfactionScore to reflect our business process.

Data Traveller

BACK TO AGENDA PAGE

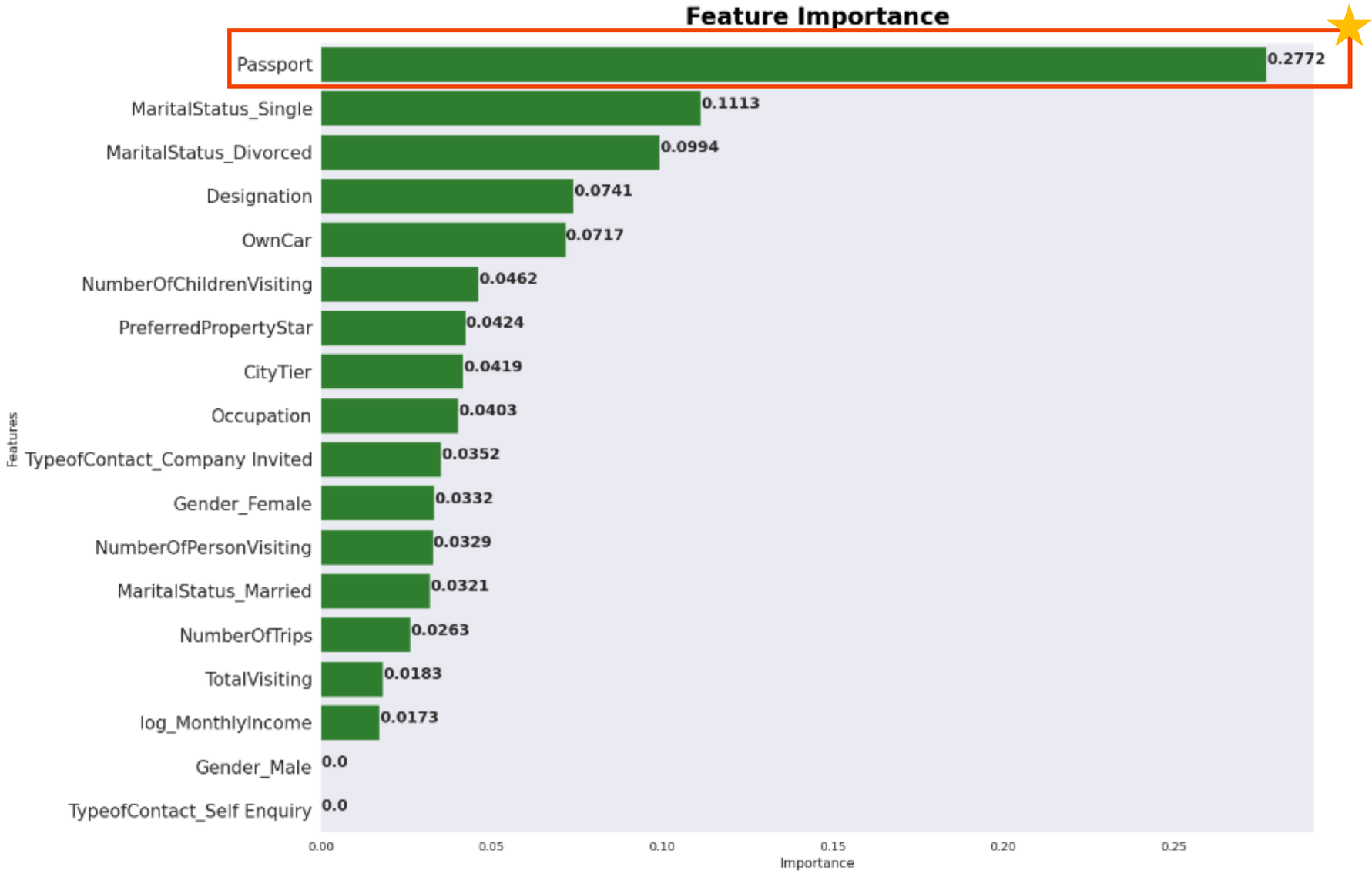
Modelling

Precision metric used to minimize incorrectly predicting customers who are not going to purchase as purchase

	Accuracy	Precision	Recall	F1	AUC
XGBClassifier	86.50	68.63	52.24	59.32	87.73
RandomForestClassifier	85.58	68.86	42.91	52.87	86.58
ExtraTreesClassifier	86.15	69.83	46.64	55.93	85.85
BaggingClassifier	84.81	63.4	45.9	53.25	86.48
GradientBoostingClassifier	83.33	57.49	44.4	50.11	82.29

Data Traveller

Feature **Importance**



BACK TO AGENDA PAGE

Data Traveller

Business **Recommendation**

Based on EDA and model's feature importance

1	Provides passport registration services to help potential customer to choose various destination offered
2	Focusing our marketing channel on social media to cover our largest customer segment who is tech savvy, 26 to 35 years old
3	Marketing campaign to target the right customer profile, any of these condition : • who is living in city tier 2 & 3 • employees • entepreneur
4	Email notification blast to inform potential customer and buyer about new promo or special package
5	Prioritize follow up to potential customer (up to 6x follow up) over non-potential customer (1x follow up) to boost marketing budget efficiency

BACK TO AGENDA PAGE

Data Traveller

Conversion Rate **Simulation**

BACK TO AGENDA PAGE

All Customer	1,422
Conversion Rate	**18%**

Before Model

	Customer
True Positive	140
False Positive	64
True Negative	1,090
False Negative	128
Conversion Rate	**68%**

TP / (TP + FP)
= 140 / (140 + 64)
= 68%

After Model

Data Traveller

BACK TO AGENDA PAGE

Cost Efficiency **Simulation**

Cost per Follow Up = IDR 10,000
Num of Follow Up 4 = Average Num of Follow Up from historical data
Num of Follow Up 6 = Max Num of Follow Up from historical data

	Customer	Num of Follow Up	Cost (IDR)
All Customer	1,422	4	56,880,000
Total			**56,880,000**

Before Model

	Customer	Num of Follow Up	Cost (IDR)
True Positive	140	6	8,400,000
False Positive	64	6	3,840,000
True Negative	1,090	1	10,900,000
False Negative	128	1	1,280,000
Total			**24,420,000**

After Model

Data Traveller

BACK TO AGENDA PAGE

Customer Acquisition Cost **Simulation**

	Total Follow Up	Total Customer	Marketing Budget (IDR)	Cost (IDR)	Marketing Budget Balance (IDR)	Customer Acquisition Cost	Additional Customer Pool
Before Model	5,688	1,422	56,880,000	56,880,000	0	40,000	0
After Model	**2,422**	1,422	56,880,000	**24,420,000**	**32,460,000**	**17,172**	**1,890**

Data Traveller

Thank **You**