PentagonTrip.com

Tour and Travel

Role

Akhmad Yuzfa S I
Data Scientist Intern

Arief Rahman H
Data Scientist Intern

Bernardus Valentino
Data Scientist Intern

Milenia Nadia A P
Data Scientist Intern

Sean Frederic W
Data Scientist Intern

Table of Contents

01 Background

Travel Package

Basic, Standard, Deluxe, Super Deluxe, and King

Problem

Inefficient telemarketing performance

Customers were contacted randomly, 82% of the customers didn't take the package

Goal

Maximize revenue by maximizing the telemarketing performance

Objectives

- Create **predictive models** to determine potential customers to buy existing travel packages based on available data

Business Metrics

Increasing revenue

Impact: a decrease in telemarketing cost to revenue ratio

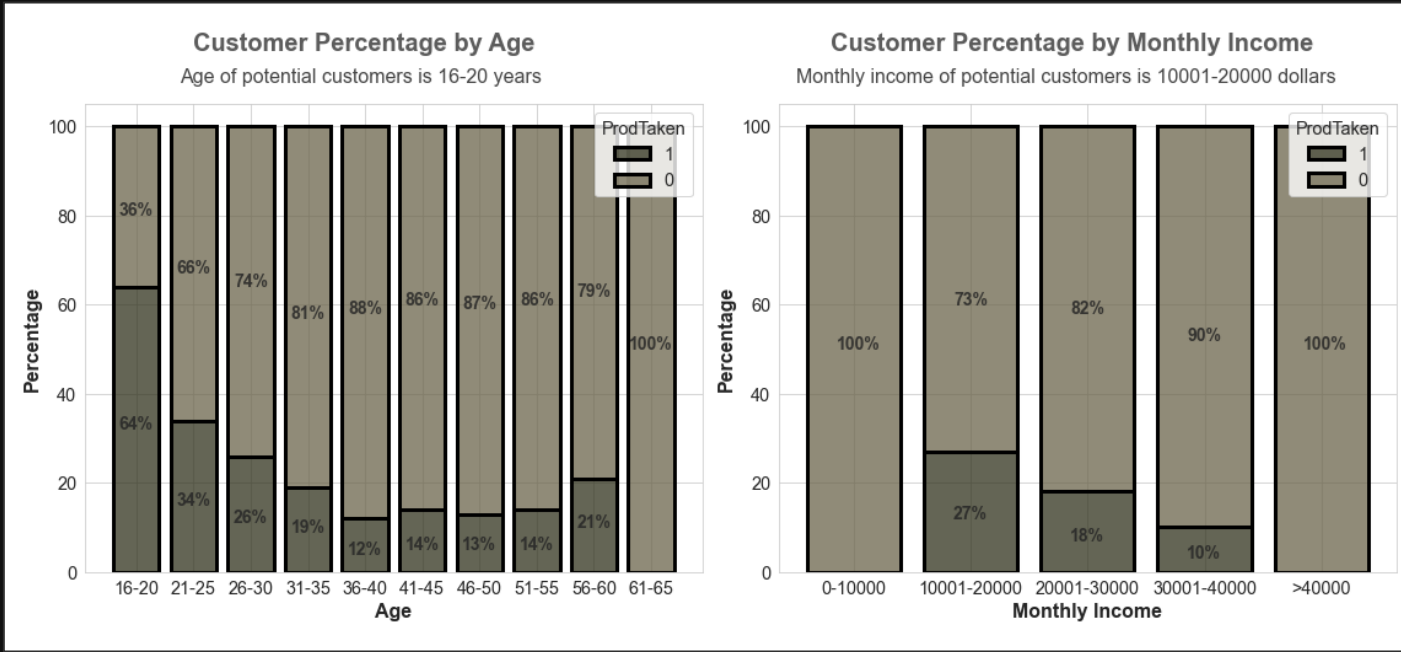
02 Dataset

04 Exploratory Data Analysis

Columns

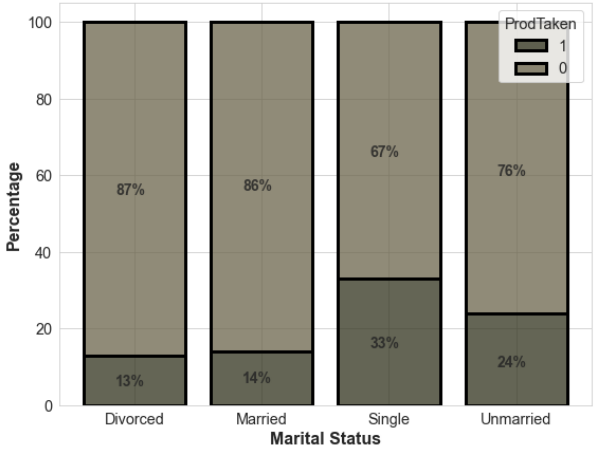
1. **ProdTaken (Target)**
2. CustomerID
3. Age
4. TypeofContact
5. CityTier
6. DurationOfPitch
7. Occupation
8. Gender
9. NumberOfPersonVisiting
10. NumberOfFollowups
11. ProductPitched
12. PreferredPropertyStar
13. MaritalStatus
14. NumberOfTrips
15. Passport
16. PitchSatisfactionScore
17. OwnCar
18. NumberOfChildrenVisiting
19. Designation
20. MonthlyIncome

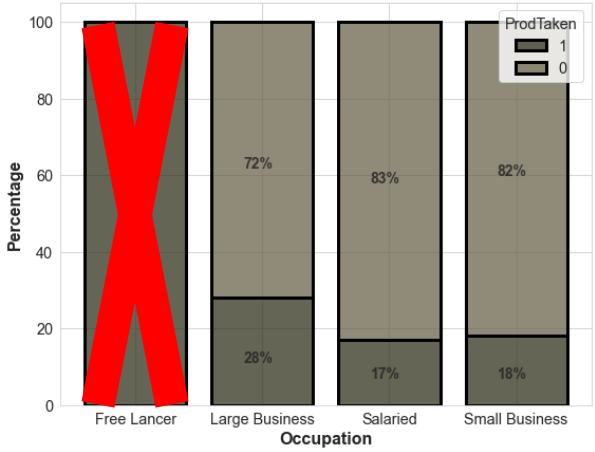
Numerical Variables

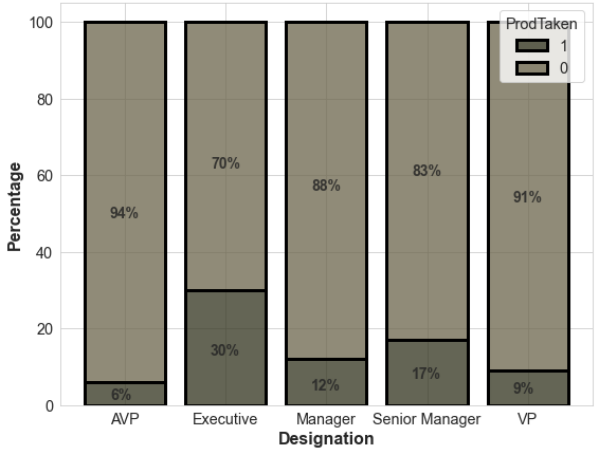


3.0 Exploratory Data Analysis

Categorical Variables







Exploratory Data Analysis

Heatmap

Top 3 features that are most correlated with the target (ProdTaken)

Passport, Age, dan MonthlyIncome

04 Preprocessing

Missing Values
- **Drop** : TypeofContact, NumberOfChildrenVisiting
- Filling with **mode**: Age, NumberOfTrips
- Filling with **median**: DurationOfPitch, NumberOfFollowups, MonthlyIncome

Outliers
Handling using **Z-Score**

Scaling
StandardScaler : NumberOfTrips, MonthlyIncome, and DurationOfPitch

Feature Encoding
- **Label Encoding**: TypeofContact, Passport, OwnCar, Gender, PreferredPropertyStar, PitchSatisfactionScore, and CityTier

- **One Hot Encoding**: Occupation, ProductPitched, MaritalStatus, and Designation

Splitting Train and Test
80 : 20

Imbalancing
Oversampling **SMOTE** with ratio **2 : 1**

Amount of Data
Initial: 4888 x 20

After pre-processing: 4787 x 33

Data training:

Target	Before SMOTE 3829 x 33	After SMOTE 4680 x 33
0	3120	3120
1	709	1560

Data testing: 958 x 33

05 Modeling

MODEL	MODEL PERFORMANCE				CONFUSION METRICS			
	ACCURACY	PRECISION	RECALL	EXECUTION TIME	PREDICT T ACTUAL T	PREDICT F ACTUAL F	PREDICT T ACTUAL F	PREDICT F ACTUAL T
Logistic Regression	0.84	0.81	0.31	0.043 s	61	745	14	138
Logistic Regression (hyperparameter tuning)	0.84	0.82	0.30	0.290 s	59	746	13	140
Decision Tree	0.90	0.78	0.71	0.021 s	141	720	39	58
Decision Tree Regularization	0.83	0.74	0.27	0.363 s	53	740	19	146
Random Forest	0.90	0.99	0.53	0.355 s	105	758	1	94
XGBoost	**0.93**	**0.95**	**0.70**	**0.219 s**	**140**	**752**	**7**	**59**
KNN	0.87	0.88	0.41	0.004 s	81	748	11	118
KNN Regularization	0.82	0.83	0.15	3.830 s	29	753	6	170

Business Metrics

Increasing Revenue

Assumptions

1. Package Price

Package	Price
Basic	**$1000**
Standard	**$2000**
Deluxe	**$3000**
Super Deluxe	**$4000**
King	**$5000**

https://costaricaexperts.com/package/best/

2. Telemarketing cost for 1 customer **$50**
https://www.magellan-solutions.com/blog/cost-of-telemarketing

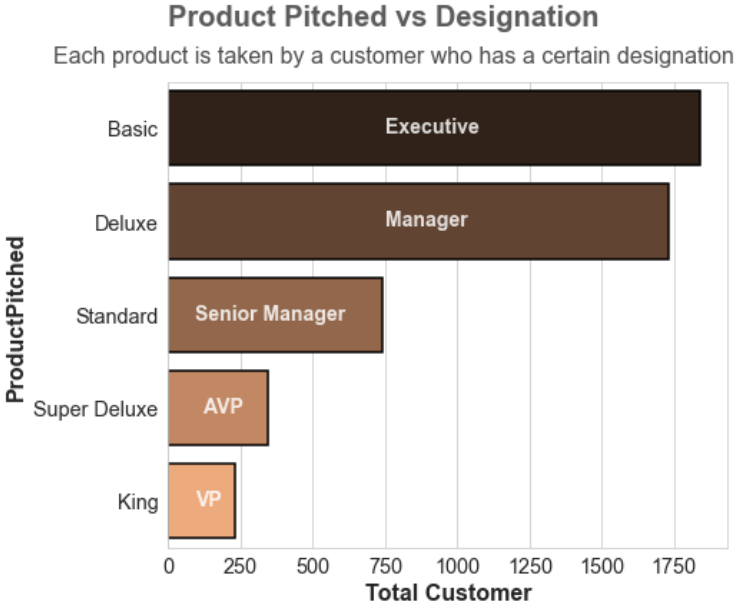
Potential Impact

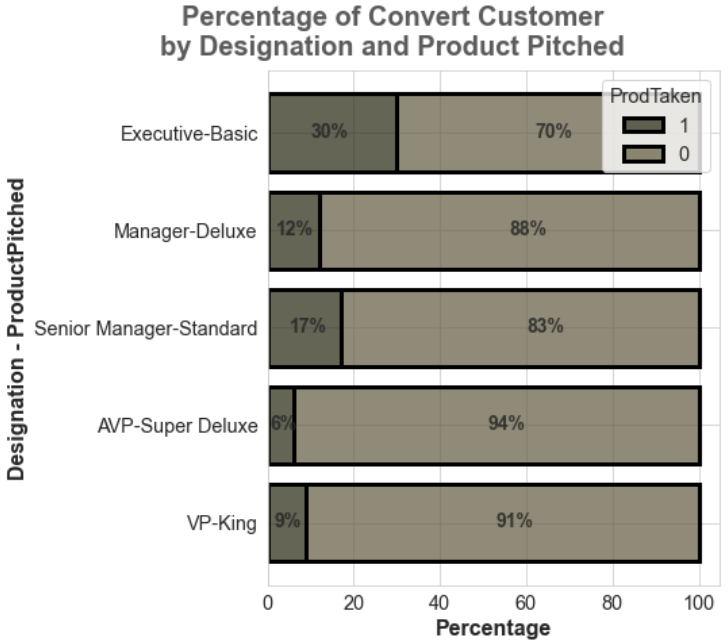
	Before Data test 958 rows	**After** Data test 958 rows (assumption: after recall calculation)	
Expected Revenue Sum(Price(i) * TotalCustomer(i))	**$2,201,000**	**$2,201,000**	**Fixed**
Telemarketing Cost DurationOfPitch*NumberOfFollowups*TeleCost	**$46,342.5**	**$46,342.5**	**Fixed**
Actual Revenue Expected Revenue * BuyingPercentage	= $2,201,000*18% **$396,180**	= $2,201,000*95% **$2,090,950**	▲ **427.8 %**
Spending Revenue on Telemarketing Cost (Telemarketing cost / Actual Revenue)*100	**11.7 %****	**2.2 %**	▼ **9.5 %**

** **Warning!**
"You should spend 2–5% of your sales revenue on marketing"
https://nuphoriq.com/create-a-marketing-budget/

7 Business Recommendation

Give the package not only based on customer designation





Customers only given specific package according to their designation

Not many customers take the product

Thank you