Nov 15th, 2022

TEXAS
The University of Texas at Austin

Trips and Travel.com customer analysis

MIS 382N: Group Project

Monica Martinez, Aakash Talathi, Neel Sheth, Yvonne Wang,
Muskaan Singhania

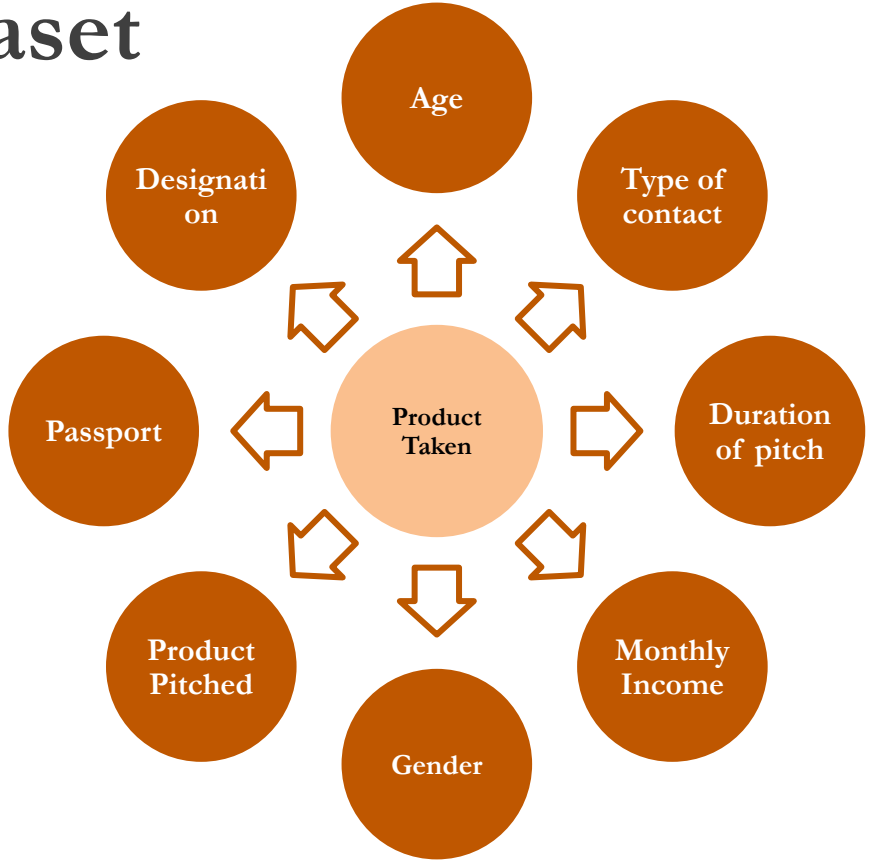
Problem Statement

Trips & Travel.Com wants to expand its customer base by adding a new product to its product offerings.

We aim to identify the potential customer base to market this product to, in order to increase the ROI on the marketing expenditure.

Understanding the Dataset

- What are the variables that affecting the target - "Product Taken"
- The dataset consists of 4888 rows with 18 features
  - Categorical
  - Quantitative



Approach

Exploratory Data Analysis

- Scatter Plots

Feature Selection

- 5 variables out of 21 variables selected

Data Preparation

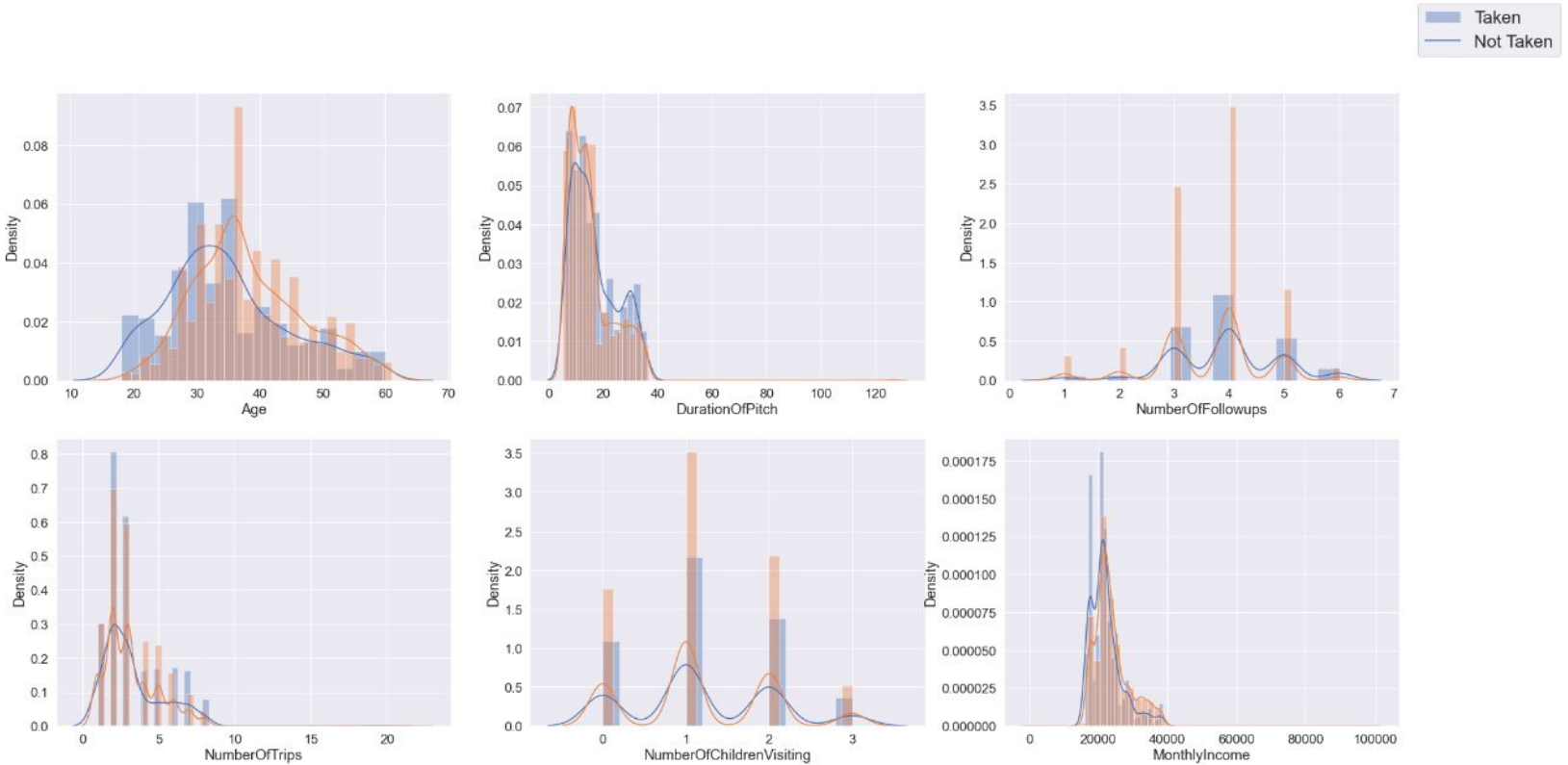
- Filtering the dataset

Modeling

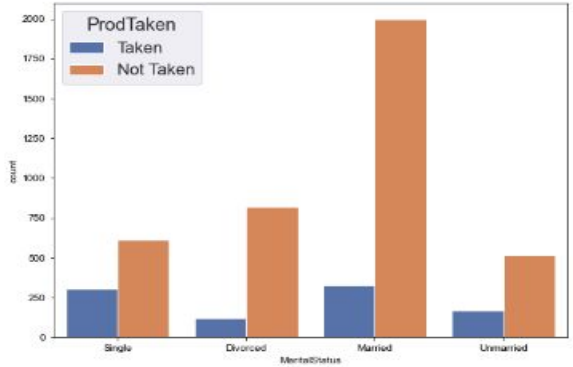
- Logistic Regression
- Customer Scoring
- Random Forests

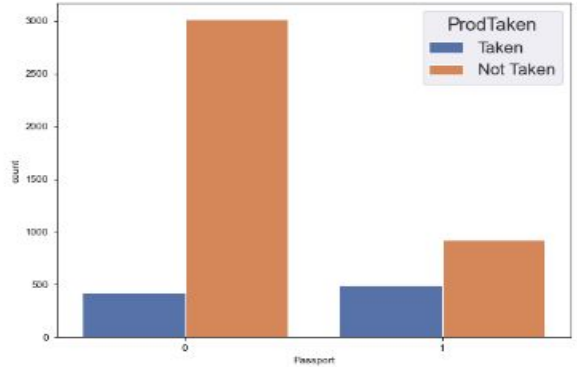
Results

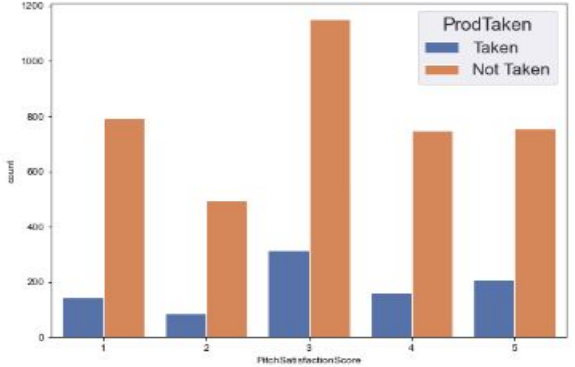
Exploratory Data Analysis - Continuous

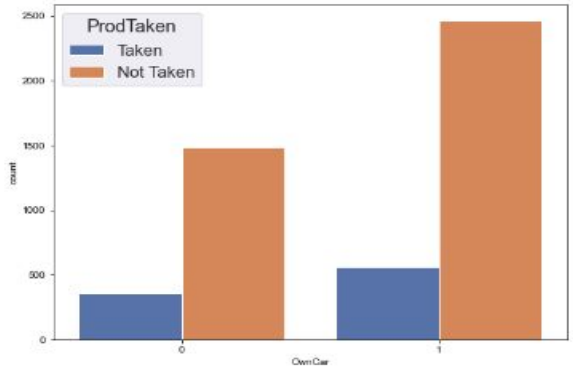


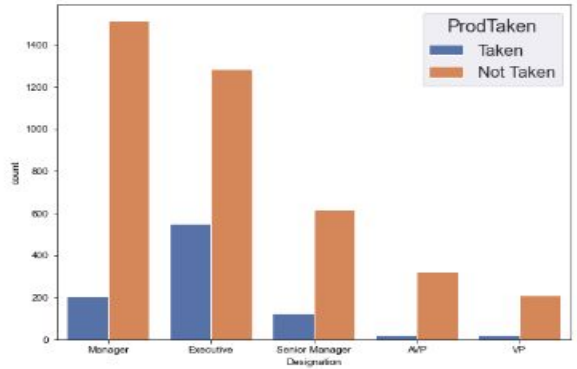
Exploratory Data Analysis - Categorical

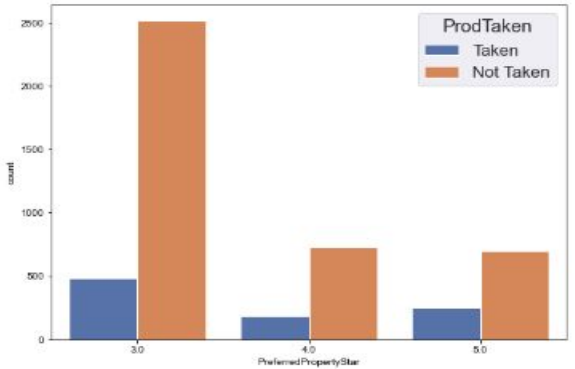












Feature Importance

- Feature ranking based on Chi2 test

- MonthlyIncome has the most impact on the target

MonthlyIncome	91649.5
Passport	235.8253
Age	222.9664
DurationOfPitch	128.6842
Marital Status	127.5178
Designation	29.22755
CityTier	18.72158
NumberOfFollowups	16.33239
Occupation	11.98029
TypeOfContact	8.497068
PreferredPropertyStar	8.452716
PitchSatisfcationScore	7.822561

Data Preparation

- Replaced missing values with median
- Categorical Columns and Target encoded as 0 & 1
- Dropped irrelevant columns

Models

- Logistic Regression
- Customer Scoring
- Random Forests

Variable Selection

- Tried 4 models - top 2 features, top 5 features, top 10 features, and all features.
- Model with top 5 features produced best results

MonthlyIncome	91649.5
Passport	235.8253
Age	222.9664
DurationOfPitch	128.6842
Marital Status	127.5178
Designation	29.22755
CityTier	18.72158
NumberOfFollowups	16.33239
Occupation	11.98029
TypeOfContact	8.497068
PreferredPropertyStar	8.452716
PitchSatisfcationScore	7.822561

XLStat Logistic Regression Coefficients

Source	Value
Intercept	0.490
Duration of Pitch	0.026
Marital Status	-0.335
Monthly Income	-0.402
Passport	1.448
Age	-0.033

Average Response Rates and Lifts

Avg Response Rate	Avg Lift
18.8216%	87.2344%

- Average Response Rate of 18.8216% tells us that there is a 18.8216% chance that a customer responded positively
- Average lift of 87.23% tells us that there is an increase of 87.23% in sales

Average Marginal Effects

Avg ME Dur	Avg ME MS	Avg ME I	Avg ME P	Avg ME A
0.3438%	-4.4053%	-5.2913%	19.0438%	-0.4376%

- Average Marginal Effect for each variable tells us the unit change in that variable for a one unit change in the response

Profit Calculation

Profits
2800
-200
2800
-200
-200
-200

ECLV	3000
Solicitatio	200
Cutoff	0.066667

- Calculated profit for each customer, based on Expected Customer Lifetime Value and Solicitation Cost to make a cutoff.
- ECLV and Solicitation Cost were taken as the average for the travel industry according to: https://blog.datumize.com/what-is-the-value-of-a-new-customer-in-the-travel-industry

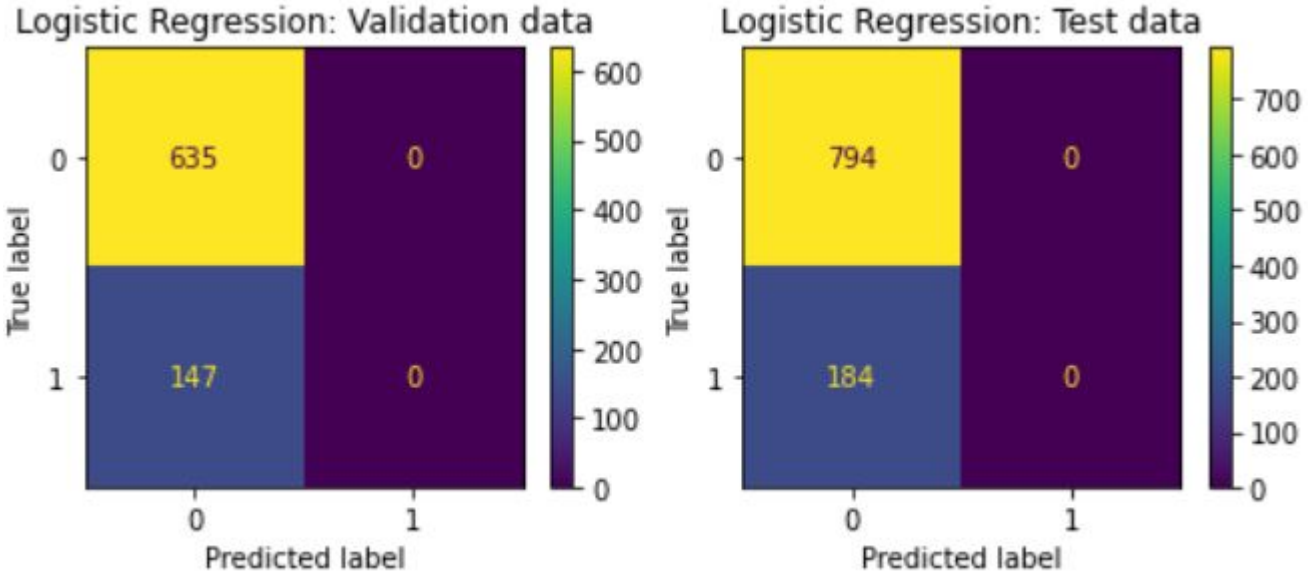
Profits with Targeting	$1,653,000.00
Profits without Targeting	$1,182,400.00

%Targeted and Profits with and without targeting

%Targeted	77.8435%
Profits with Targeting	$1,653,000.00
Profits without Targeting	$1,182,400.00

- Percent of customers targeted - 77.84%
- Difference in profit of approximately $471,000.

Logistic Regression



Accuracy Score:
0.8289641943734015

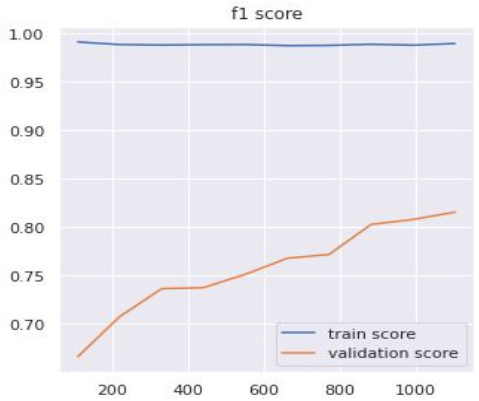
Random Forest Classifier

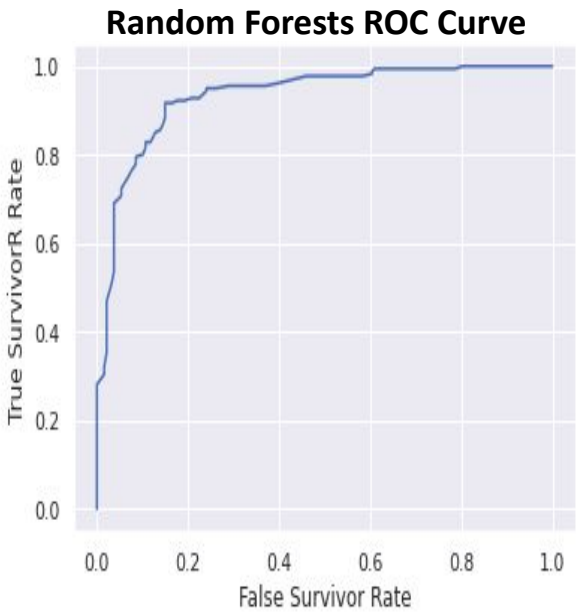
```
Accuracy =  0.8559782608695652
-
[[162  25]
 [ 28 153]]
-
              precision    recall  f1-score   support

           0       0.85      0.87      0.86       187
           1       0.86      0.85      0.85       181

    accuracy                           0.86       368
   macro avg       0.86      0.86      0.86       368
weighted avg       0.86      0.86      0.86       368

-
```





Results

- By Customer Scoring we are able predict that 18% of the response rate is positive, which we think is pretty good, considering most people's negative response to a sales pitch.
- From the Logistic Regression model we can predict future customers from the featured selection with 82% accuracy.
- We were further able to improve this accuracy to 86% with a Random Forest Classifier.
- By targeting specific customers, we can improve profit by almost half a million dollars.