Visit with us

Travel Package Purchase Prediction

Contents

- Business Overview of the problem and solution approach
- Key findings and insights which can drive business decisions
- Model overview and performance summary
- Business recommendations

Business Overview of the problem and solution approach

- You are a Data Scientist for a tourism company named "Visit with us". The Policy Maker of the company wants to enable and establish a viable business model to expand the customer base.

- A viable business model is a central concept that helps you to understand the existing ways of doing the business and how to change the ways for the benefit of the tourism sector.

- One of the ways to expand the customer base is to introduce a new offering of packages.

Business Overview of the problem and solution approach

- Currently, there are 5 types of packages the company is offering - Basic, Standard, Deluxe, Super Deluxe, King. Looking at the data of the last year, we observed that 18% of the customers purchased the packages.

- However, the marketing cost was quite high because customers were contacted at random without looking at the available information.

- The company is now planning to launch a new product i.e. Wellness Tourism Package. Wellness Tourism is defined as Travel that allows the traveler to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being.

Business Overview of the problem and solution approach

- However, this time company wants to harness the available data of existing and potential customers to make the marketing expenditure more efficient.

- You as a Data Scientist at "Visit with us" travel company must analyze the customers' data and information to provide recommendations to the Policy Maker and Marketing Team and also build a model to predict the potential customer who is going to purchase the newly introduced travel package. (Classification)

- **Objective**: To predict which customer is more likely to purchase the newly introduced travel package.

Business Overview of the problem and solution approach

Target variable:`ProdTaken`.

What are we doing?

Creating an inferential model on the training data from a sample of this dataset. Then using the same model to predict the likelihood that a customer will purchase a package using the `ProdTaken` feature. We will look at the model's accuracy (and other metrics) to gauge the reliability of the prediction in applying it to a marketing campaign. We will give insights from analysis and recommendations using the business context.

What variables explain the change in the target variable?

We will look at each variable individually for it's validity to be used in the an inferential model. Then we will prune where necessary or allow the hyperparameter tuning to decide for us based on the type of algorithm we use.

Contents

- Business Overview of the problem and solution approach
- Key findings and insights which can drive business decisions
- Model overview and performance summary
- Business recommendations

Key findings and insights which can drive business decisions

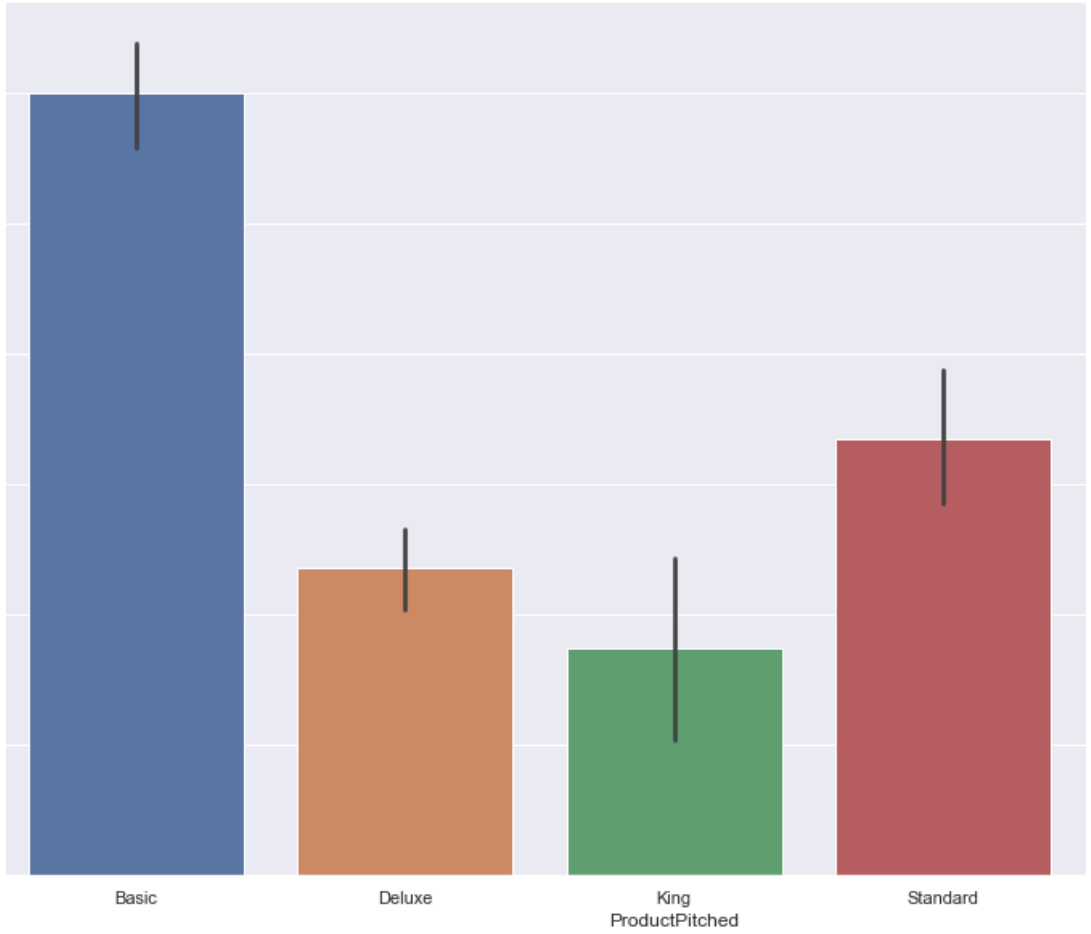
- Data includes:
  - Survey data with customer responses to travel package purchases, passport, etc.
  - Demographic information disclosed from survey, i.e.: marital status, gender, etc.
  - Travel meta-data: # of trips, # of children, etc.
  - Sales interaction information: i.e., satisfaction with pitch, # of follow-ups, etc.

Key findings and insights which can drive business decisions

Customer purchases are focused on the Basic package.

- More people took the Basic over the Deluxe and King combined.

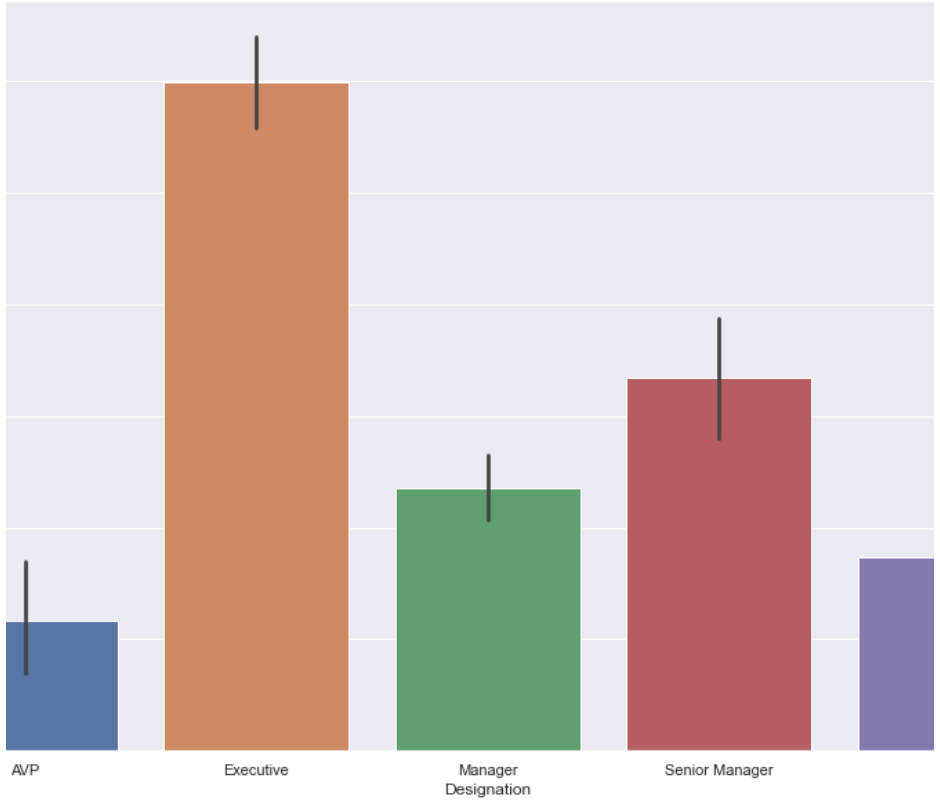
- -Standard is the second most chosen.



Key findings and insights which can drive business decisions

Executives chose to purchase a package.

- 552 of the 920 purchasers were exec's
- That's more than all the other categories combined.



Key findings and insights which can drive business decisions

ProdTaken 0 1 All	0	1	All
Divorced	824	124	948
Married	2012	326	2338
Single	610	302	912
Unmarried	514	166	680
All	3960	918	4878

Married people bought packages more than other status who did.

- Single people are more likely to purchase packages than married people

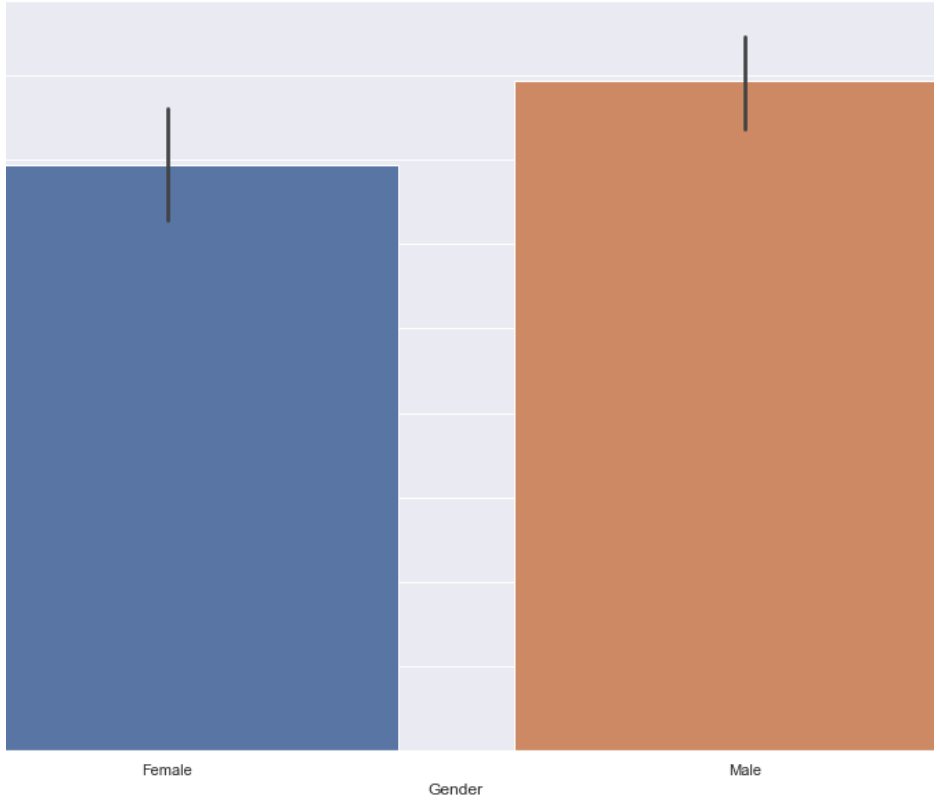
- Married people have the worst % of purchasing to non-purchasers.

Key findings and insights which can drive business decisions

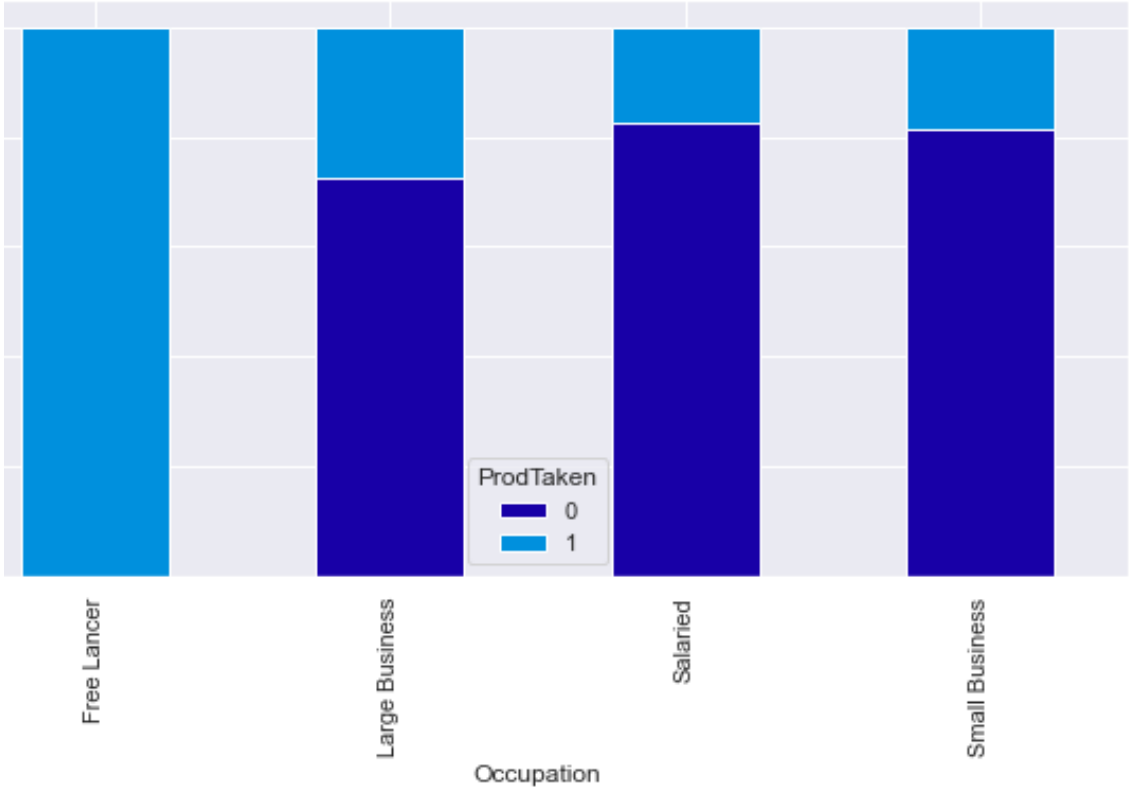
Men buy packages.

- Only 18% of women did.

- More than half of all purchasers were men.



Key findings and insights which can drive business decisions



Freelancers buy packages (2 customers only – can we get more data?).

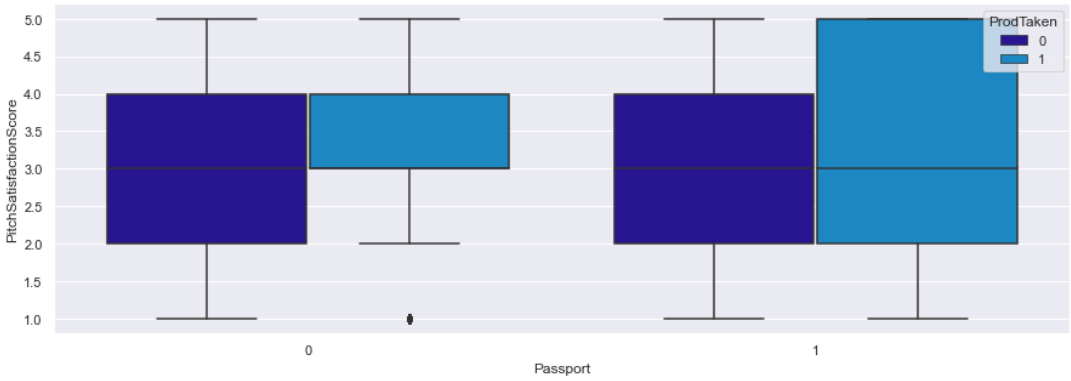
- Of those who did purchase, most were Salaried.
- About 1/3 large business customers purchased packages

Key findings and insights which can drive business decisions

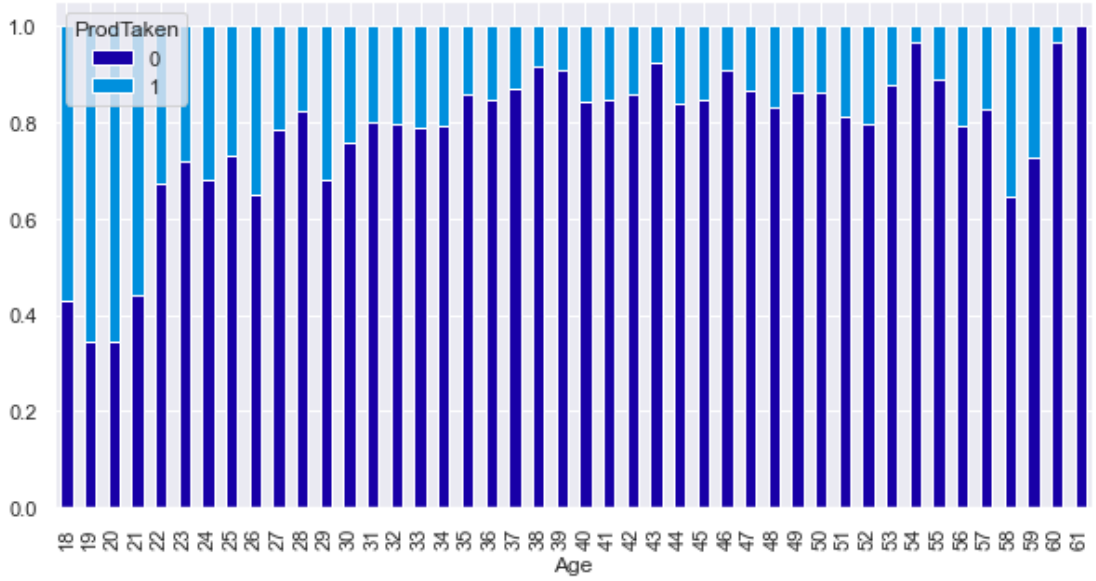
People with passports bought packages.

- Only ~23% of passport owners did *not* buy a package.

- 53% of purchasers are passport holders



Key findings and insights which can drive business decisions



Age plays a role in purchasing.

- Younger people 19 – 32 tend to purchase packages.

- 36-year-olds bought the most packages.

Contents

Model overview and performance summary

- Modeling was done using Classifiers for a binary problem: Did a customer purchase a travel package or not
- Models were built using different types of classifiers: Decision Trees, Random Forest, Bagging, Boosting and Stacking.
- Models were all scored using the same metrics: accuracy, precision, recall and a confusion matrix.

Model overview and performance summary

- Given the business problem of correctly identifying those customers who did (and by assumption who *will*) purchase travel packages our target score was **recall.**

- Recall is used to identify sensitivity to False Negatives. The higher the recall score the better.

- The Confusion Matrix is used to identify the % of misclassification of purchasers to non-purchasers using False Negative or Type II error classification.

Model overview and performance summary

All models were compared.

The models with 100% recall were over-simplified models that are not useful.

The chosen best model was the Cost-Complexity Pruned Decision Tree for it's balanced metrics.

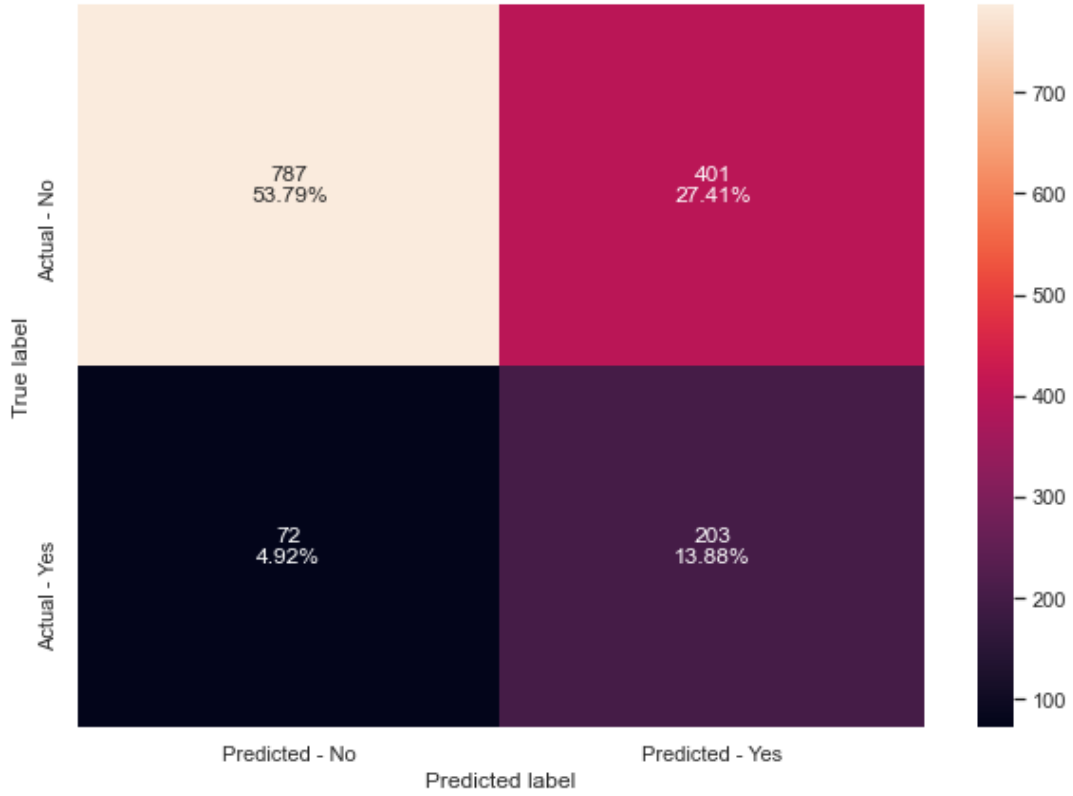
Model overview and performance summary

- We can predict within 68-76% accuracy the best fit customers for purchasing a travel package.
- We are sensitive to 74-96% of the variance in the decision.
- Our error range from the same model for falsely predicting customers who are not purchasers is ~5%.

Model overview and performance summary

The False Negatives were reduced to ~5% with this decision tree model.

This gives us a decent safety range for mis-classifying purchasers.



Model overview and performance summary

- Features of the data for predicting or correlating the most likely purchasers of travel packages are customer's monthly income, their age and the customer's passport holding status.

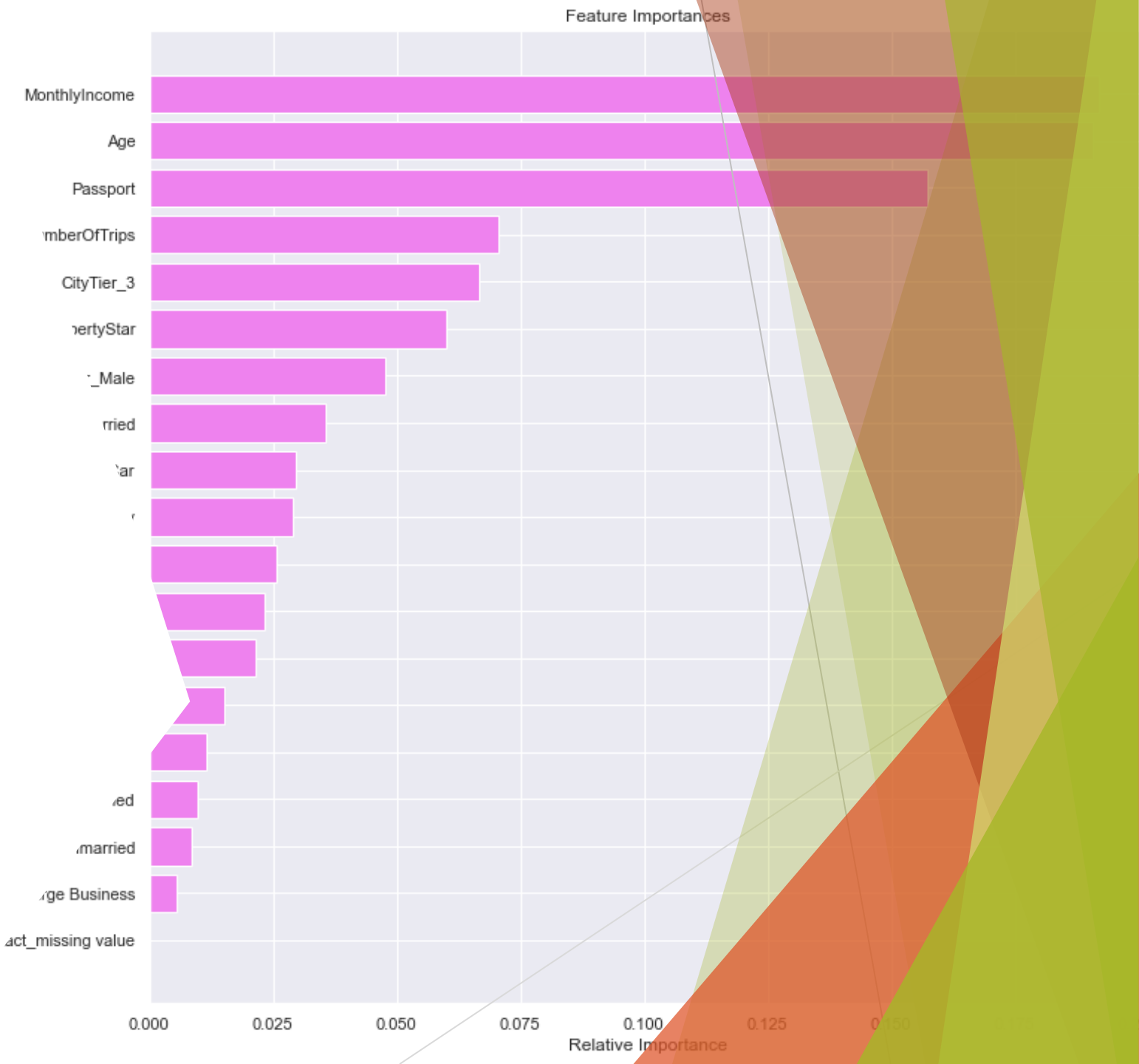
- These were drawn from a Decision Tree model that was pruned using a method called Cost-Complexity.

- This information is useful when making decisions about which factors to weigh in targeting a customer base.

Features of most importance

Of least importance are:

- Large Business
- Unmarried
- Salaried



Contents

- Business Overview of the problem and solution approach
- Key findings and insights which can drive business decisions
- Model overview and performance summary
- Business recommendations

Business recommendations

1. What have we done well in the past and what can we predict on?

- Target marketing to the population sampled and mentioned in the market segmentation. Focus on the important features: income, age and passport holding.
- Bring the customer interaction data from EDA to the sales team to discuss the pros-and-cons of what we found in EDA based on what is possible for them. (do cost-benefit analysis on product lines).
- Not worrying about Designation because the titles across companies/industries are varied in meaning and so the cost of focusing on this variable could be put into other variables.

Business recommendations

2. What are opportunities we are missing out on?

- Married and older people who could get some packages with special offerings, coupons or discounts.

- Targeting more freelancers to get their data into the mix.

- Capturing all the available data with data-entry or survey-entry errors or missingness.

Business recommendations: Market Segmetation

ProdTaken : Whether the customer has purchased a package or not in the past is our determining factor. All future ones from the model are ones who have in the past.

Age : 29 - 36 years old

TypeofContact : Self Inquiry

CityTier : Tier 1

Occupation : Salaried

Gender : Male

NumberOfPersonVisiting : 3

PreferredPropertyStar : 3

MaritalStatus : Single

NumberOfTrips : 2

Passport : 1: Yes

OwnCar : 1: Yes

NumberOfChildrenVisiting : 1

Designation : Executive

MonthlyIncome : Gross monthly around 22,178.89 USD ($).

PitchSatisfactionScore : 3

ProductPitched : Basic

NumberOfFollowups : 4

DurationOfPitch : 8 to 9minutes

Further Improvement

Do misclassification analysis on the segments of the customers that we are not confident in predicting. We can build more heuristic rules about these customers so targeting is more accurate. To say that the model is not so great at predicting on certain profiles for customers.

External data: Try to investigate what other external factors could influence the purchase of travel packages (i.e., season, weather, major events…)

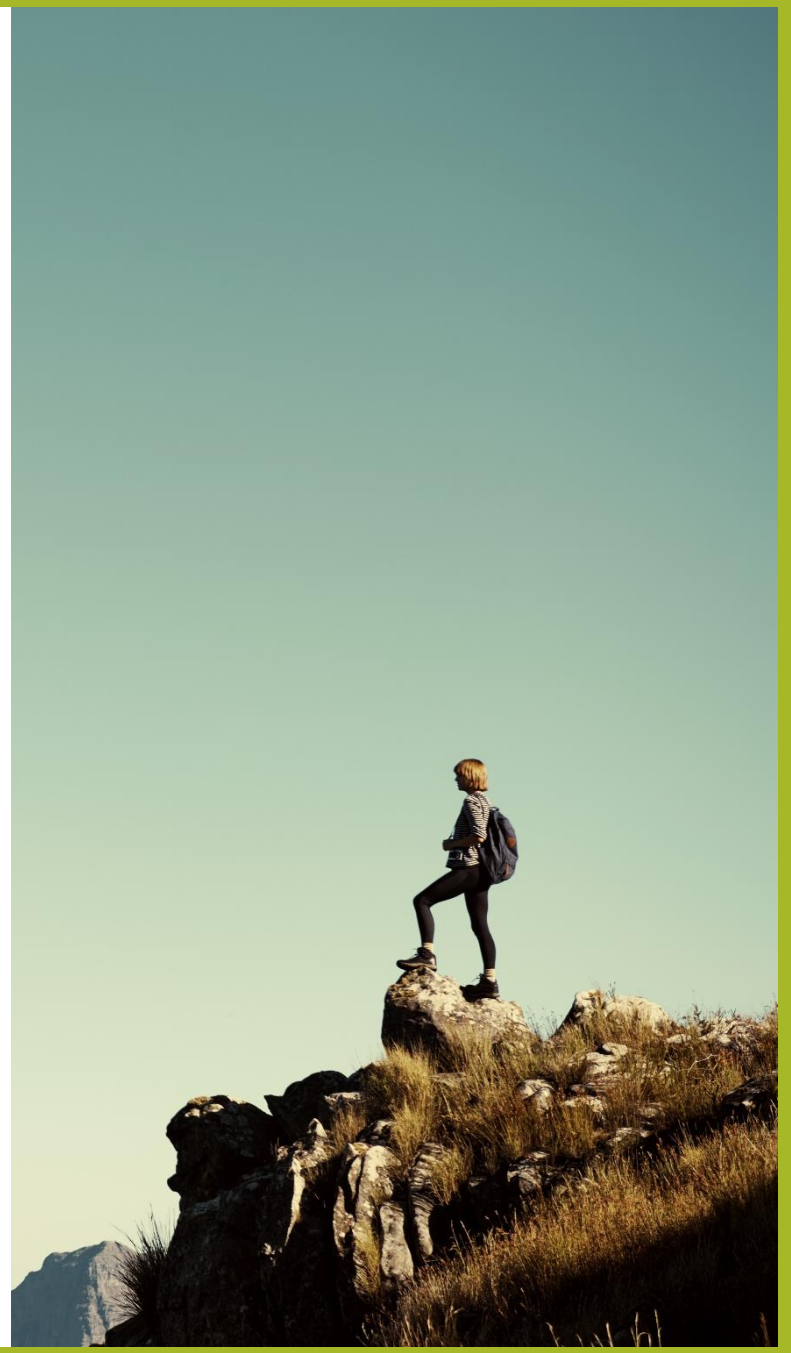
Garner additional data through a customer survey asking about travelers attitudes to maintain, enhance or kick-start a healthy lifestyle, and support or increase one's sense of well-being.

Pushing a Package

Basic

Wellness

King



Thank You

josephmbalog@gmail.com