Travel With Us

Contents

- Business Problem Overview and Solution Approach

- Data Overview

- Exploratory Data Analysis (EDA)

- Key Insights from EDA

- Customer Profile based on Packages

- Model Evaluation Criterion

- Model Performance - Bagging

- Model Performance - Boosting

- Model Comparison

- Business Insights & Recommendations

Business Problem Overview and Solution Approach

- 'Visit With Us' is a tourism company that currently offers five types of packages to it's customers, namely Basic, Standard, Deluxe, Super Deluxe and King. The Policy Makers of the company wants to establish a viable business model to expand the customer base by introducing a new travel package, the 'Wellness Tourism Package'

- Build Ensemble Technique ML models using the data of the existinn customers

- Use the models to target potential customers (including new customers) who are more likely to purchase the new package

Data Overview

- The data contains information of 4888 customers

- Customer information include age, gender, occupation, designation, marital status, monthly income, city tier, whether or not the customer owns a car, passport and also if the customer purchased the package in the previous year

- Customer interaction data with the salesperson

- Fixed the data entry error in Gender column

- Imputed missing values in columns Age, Type of contact, Preferred property star, Number of children visiting and Monthly income

- Excluded the customer interaction data while model building as this information will not be available for potential new customers

Exploratory Data Analysis

- Customers with passport are more likely to buy a package

- Single and Unmarried customers tend to buy the product more

- Customers who bought the package are higher in number from CityTier 2 and CityTier 3
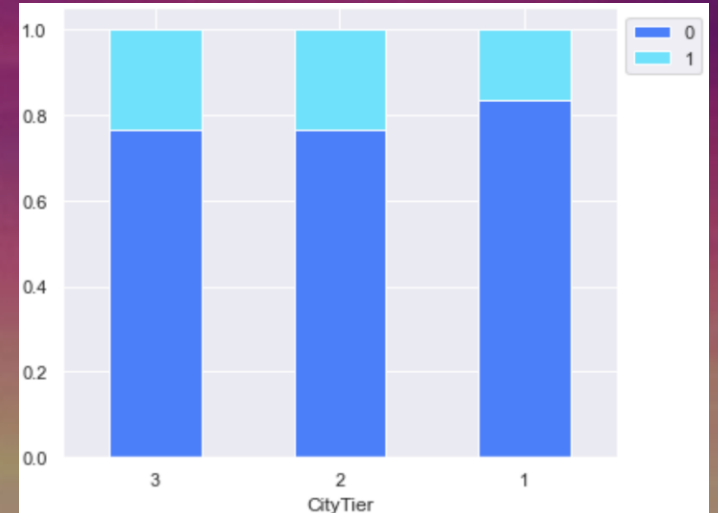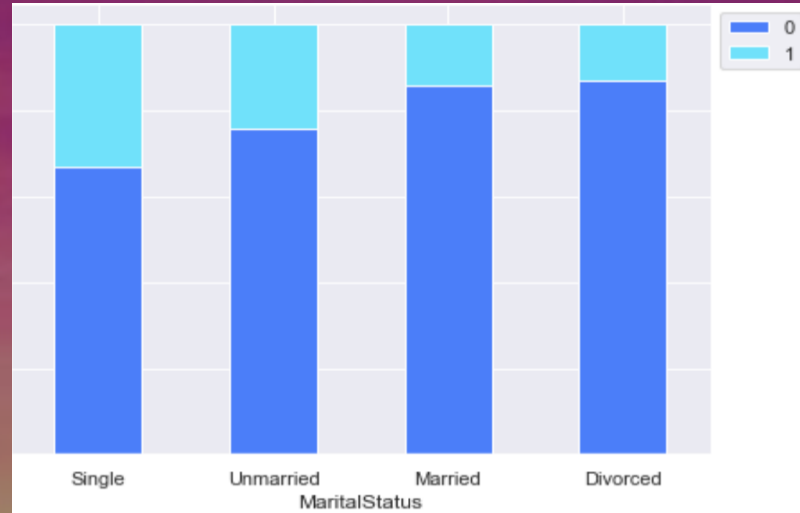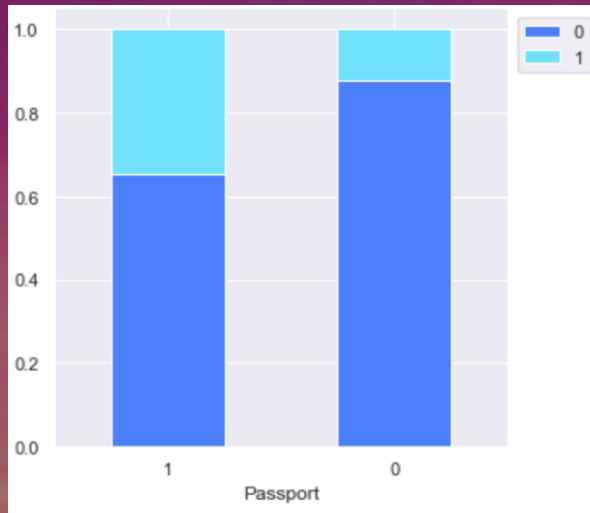
- Large business customers are more likely to buy the package

- Customers who buy the product are largely designated as Executives

- Customers who bought the package mostly prefer properties with 3 star and above
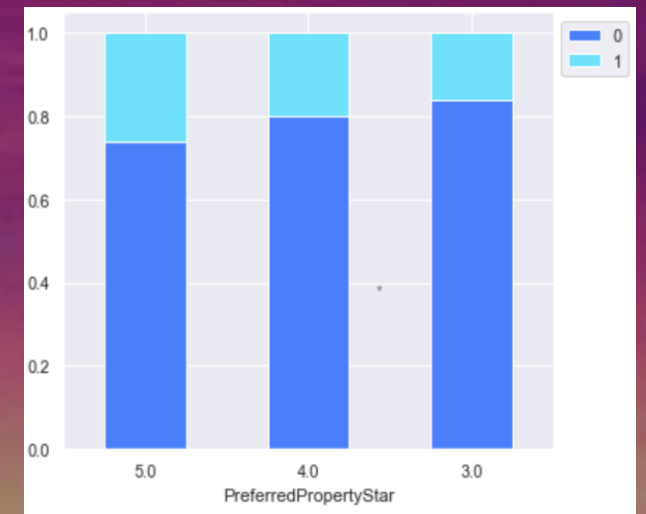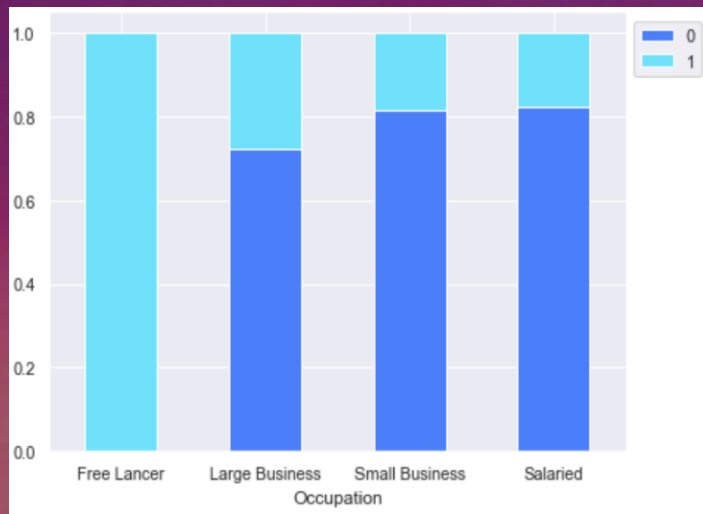
Key Insights from EDA

Customers more interested in purchasing the travel package are :

- aged between the late 20s and early 40s

- travel in a group of 2 to 4 which includes one child (on average)

- from Tier2 and Tier3 cities

- own a passport

- mostly Executives and Managers

- Female customers with a higher designation seem to be interested in the 'King' package, whereas Male customers show interest in the 'Super Deluxe'

- Single, unmarried and divorced customers highly prefer 'Basic' package while married customers prefer more expensive packages, comparatively

- who were followed up three times and above with a higher duration of pitch

- On average, customers who tend to travel three times and above, annually

- customers with an average monthly income ranging between 20K - 25K opt opt for Basic and Standard packages, 23K - 30K opt for Deluxe and Super Deluxe packages and above 33K prefer the King package

- Gender, Number of children visiting, and Own car - these factors do not add any significance to the chances of the product being purchased

Customer Profile based on Packages

Basic	Standard	Deluxe
Early 20s to late 30s	Early 30s to late 40s	Early 30s to late 50s
Self-Enquiry	Self-Enquiry	Self-Enquiry only
Salaried, Executives	Small business	Small business
CityTier1	Salaried, Senior Managers	Managers
Male, Single	CityTier3	CityTier3
2 to 3 persons, 1 child avg	Male, Married	Male, Married
50% have passport and car	2 to 3 persons, 1 child avg	3 persons, 1 child avg
3 star property & above	50% do not have passport / car	50% do not have passport
1 to 20 trips a year	3 star property & above	75% own a car
Avg 20K per month	1 to 8 trips a year	3 star property & above
	Avg 26K per month	1 to 8 trips a year
		Avg 23K per month

Model Evaluation Criterion

Model can make mistakes such as :

- Predicting a customer will purchase a travel package but in reality the customer does not purchase one.

- Predicting a customer will not purchase a travel package but in reality the customer will purchase the travel package.

Prediction of concern :

- The second prediction is our major concern as the 'Visit With Us' travel company plans to launch a new tourism package and wants to harness the available data to make the marketing expenditure more efficient. In order to do so, mistakes in the second prediction (i.e. False negatives) have to be considerably reduced.

How to reduce False Negatives :

- **Recall** score should be maximized. Greater the Recall score, higher the chances of predicting the potential customers who may purchase the new travel package.

Model Performance - Bagging

Metric / Classifier	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision	Train F1	Test F1
Decision Tree	1.0	0.873892	1.0	0.677536	1.0	0.660777	1.0	0.669052
Tuned dtree	0.748027	0.751193	0.681677	0.702899	0.400547	0.406709	0.504598	0.515272
Bagging	0.992400	0.901159	0.962733	0.550725	0.996785	0.878613	0.979463	0.677060
Bagging with tunedDecision Tree as base	0.802105	0.796864	0.673913	0.637681	0.481687	0.470588	0.561812	0.541538
Tuned Bagging Classifier	0.486407	0.475801	0.889752	0.923913	0.253652	0.254237	0.394764	0.398749
Random Forest	1.0	0.892297	1.0	0.478261	1.0	0.904110	1.0	0.625592
Tuned Random Forest	0.890968	0.867757	0.476708	0.380435	0.895044	0.820312	0.622087	0.519802

- Tuned Bagging Classifier gives the highest Recall in the test set

- The business may choose tuned Decision Tree for a good Recall with a better Precision score

Model Performance - Boosting

Metric / Classifier	Train Accuracy	Test Accuracy	Train Recall	Test Recall	Train Precision	Test Precision	Train F1	Test F1
AdaBoost	0.841859	0.837764	0.270186	0.278986	0.710204	0.663793	0.391451	0.392857
Tuned AdaBoost	0.972523	0.871847	0.885093	0.565217	0.966102	0.696429	0.923825	0.624000
Gradient Boost	0.879275	0.862986	0.428571	0.358696	0.859813	0.804878	0.572021	0.496241
Tuned Gradient Boost	0.903829	0.876619	0.546584	0.449275	0.904884	0.810458	0.681510	0.578089
XGBoost	0.991662	0.901159	0.987578	0.597826	1.000000	0.829146	0.993750	0.694737
Tuned XGBoost	0.918737	0.846626	0.958075	0.793478	0.738780	0.565891	0.738780	0.660633
Stacking	0.933061	0.829584	0.996894	0.858696	0.738780	0.529018	0.848645	0.654696

- AdaBoost classifier and Gradient Boost Classifier are most generalized models but, they perform poorly in terms of Recall

- XGBoost and Stacking models are little overfit but they are giving the highest Recall scores in the test set

Business Insights & Recommendations

- The business can use this predictive model to

  - identify potential customers who may purchase the travel packages

  - potential new customers who may purchase the packages that are offered / packages that are newly launched

  - the features that drive the customer to buy the package

- Features that impact Product taken are

  o Passport

  o Designation

  o Marital Status

  o City Tier

  o Monthly Income

  o Age

  o Number of trips annually

- customers who own a passport show more interesting in buying the product

- customers with Designation Executive, Marital Status single and City Tier 3 should be our target customers

- customers with Monthly Income 15K to 25K, Age 25 to 40 show more interest in buying a travel package

- larger the number of trips taken by a customer annually, higher is the chances of customer buying the package

- The marketing team should focus on
  - higher duration of pitch by salesperson with the customer
  - do multiple followups with the customer
  - encourage customers to get passport
  - market and company invite customers for 'King' package

- Once the 'Wellness Package' is launched, the business can collect data on customer information, their preference, product satisfaction and customer interaction so as to enable data analysis for better results

Thank You