Project 5:
Travel Package Purchase Prediction

Brandy Murray

Contents

Business Problem Overview and Solution Approach

- Core business idea
  - Visit with Us wants to expand their customer base by marketing a new Wellness Package.
- Problem to tackle
  - Increasing the number of customers who purchase packages and predict who will buy the new Wellness Package.
- Financial implications
  - If we do not analyze the data correct, and model whether the wellness package will be built or not given the variables we were given, will allow competitors a chance to move in on our customers reducing revenue and customer volume.
- How to use ML model to solve the problem
  - As the Data Scientist I will analyze the customers' data and information 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.

Data Overview

- Originally there was 4,888 rows and 20 columns
- I decided that CustomerID was not needed and ended up with 19 columns.
- The Gender column needed to have a few ‘Fe Male’ entries changed to ‘Female’.
- There were 8 columns that has missing values. Because none of the missing values were exceedingly large, I chose to replace missing values with the mean or median.
- I noticed in the Occupation column there were only 2 Free Lancer values. I changed these 2 entries to Small Business.
- At the end, I ended up with 8 categorical variables and 11 numerical variables.

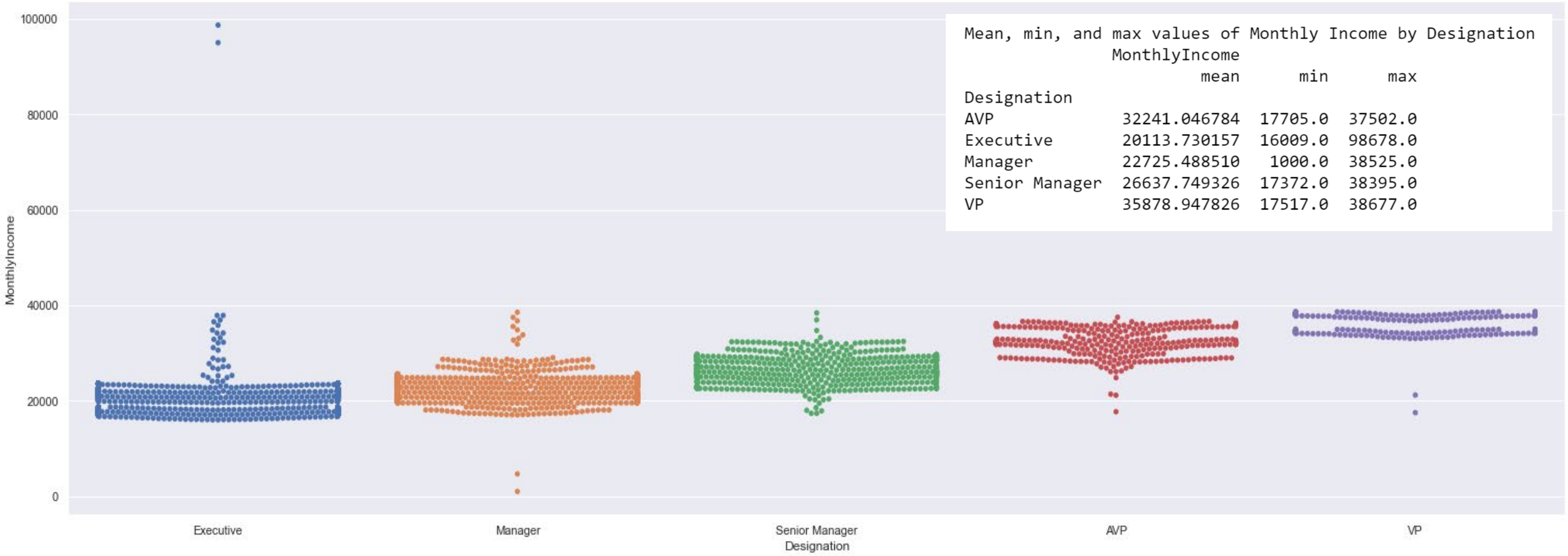
EDA

- To start the customer profiling I chose to look at the Designation of the Customers.

This visual shows us the distribution of age across the 5 Designations.

EDA

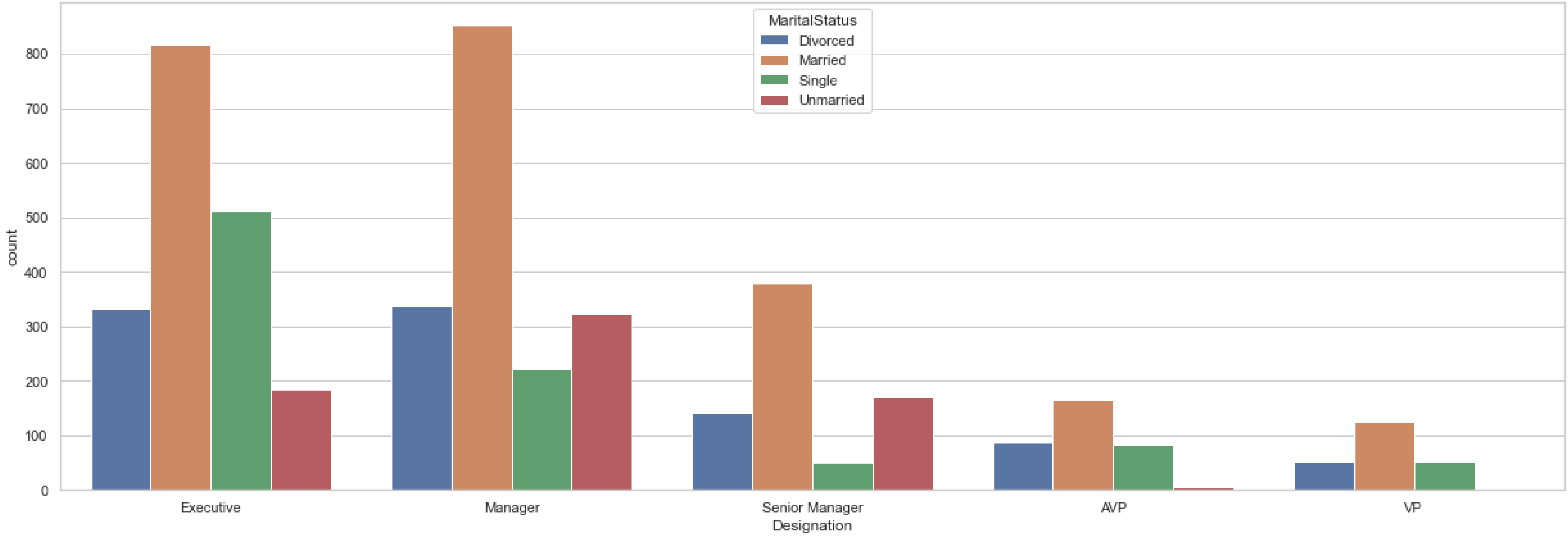
- Here you can see the monthly income across the 5 Designations.



Without concrete proof the two upper outliers for Executive were incorrect, I left them as is.

EDA

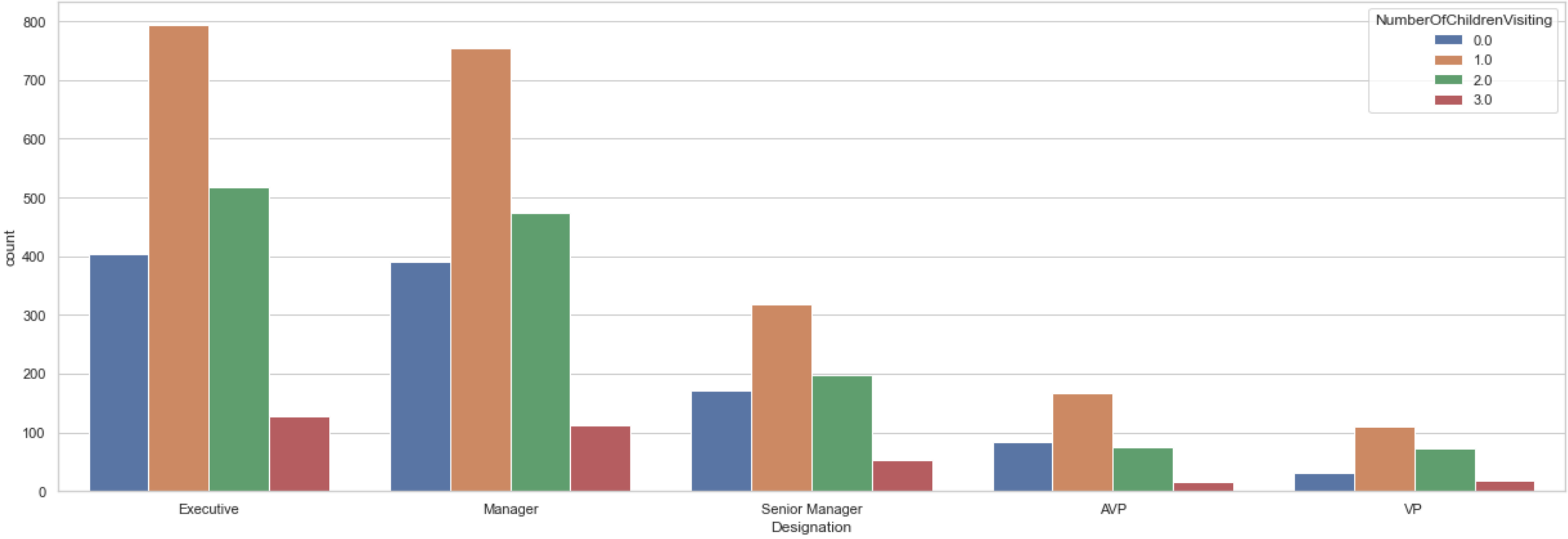
- Here you can see the marital status across the 5 Designations.



Here we can see that in each group, married is the highest marital status.

EDA

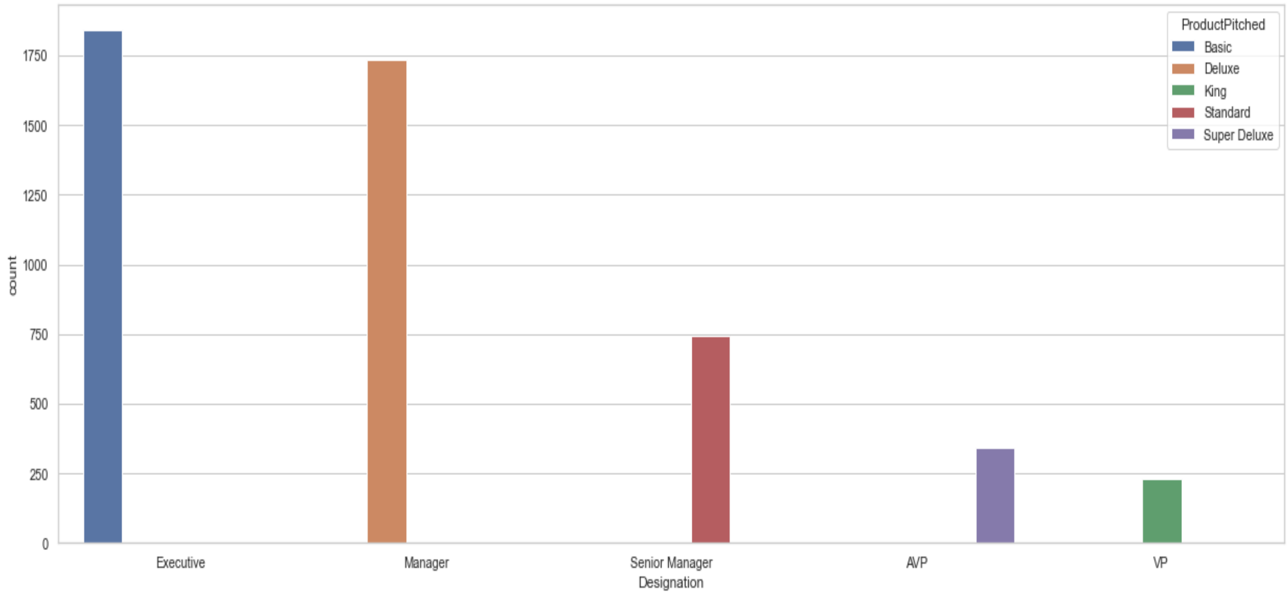
- Here you can see the number of children who came with the customer across the 5 Designations.



Here we can see that in each group, 1 child was most common.

EDA

- This shows that each Designation was only offered a specific product.

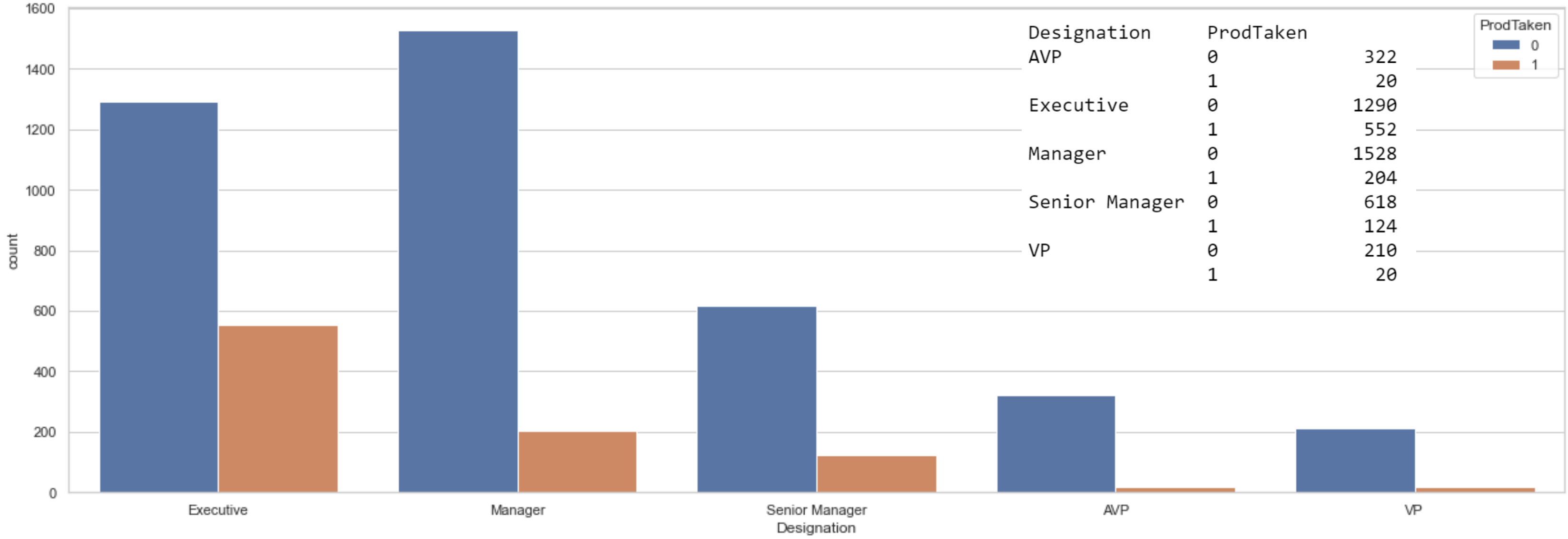


Designation	ProductPitched	
AVP	Basic	0
	Deluxe	0
	King	0
	Standard	0
	Super Deluxe	342
Executive	Basic	1842
	Deluxe	0
	King	0
	Standard	0
	Super Deluxe	0
Manager	Basic	0
	Deluxe	1732
	King	0
	Standard	0
	Super Deluxe	0
Senior Manager	Basic	0
	Deluxe	0
	King	0
	Standard	742
	Super Deluxe	0
VP	Basic	0
	Deluxe	0
	King	230
	Standard	0
	Super Deluxe	0

If we remember the income between each Designation was not that different. Why were they only offered one product?

EDA

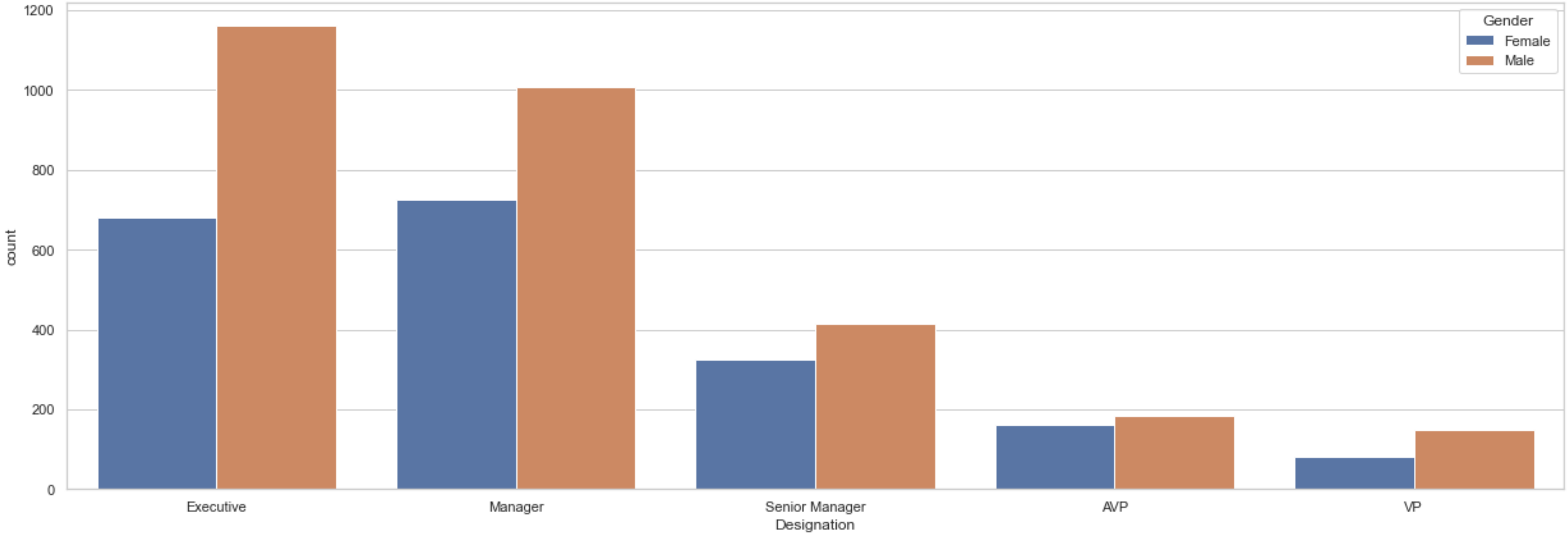
- This shows that in every Designation the Product was not taken.



Without knowing more information about the prices or what only one product was chosen is hard to interpret why the turn down rate is so high.

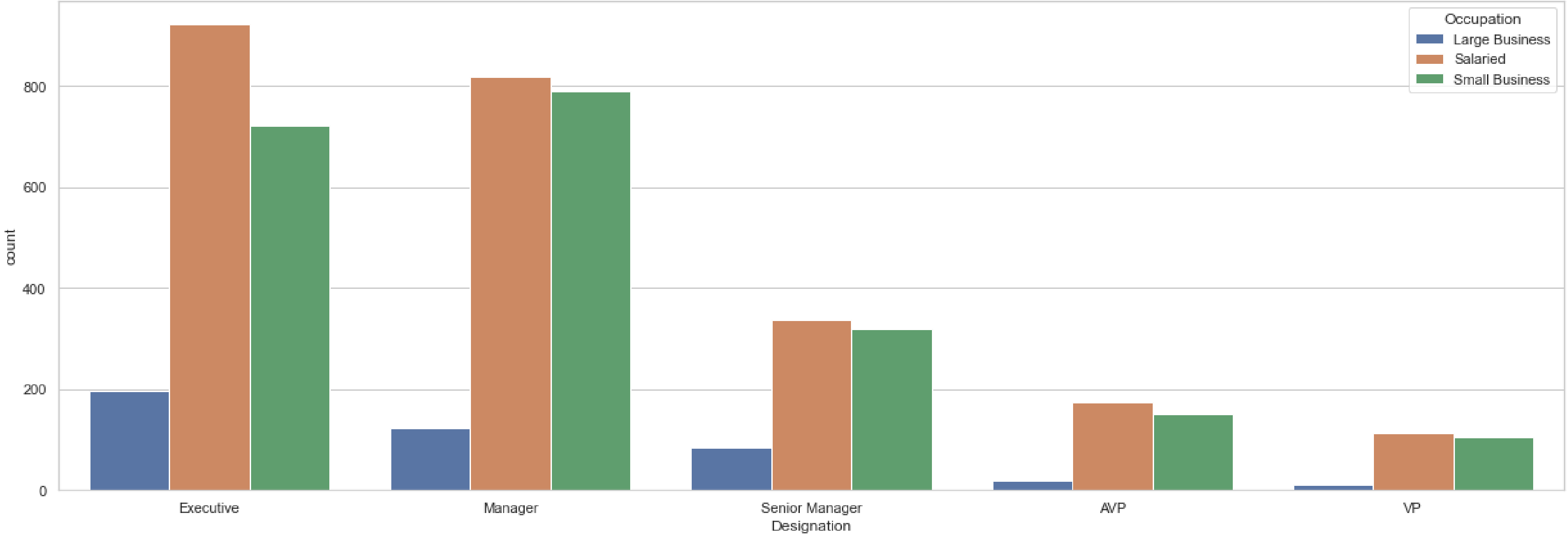
EDA

- The majority of customers were male.



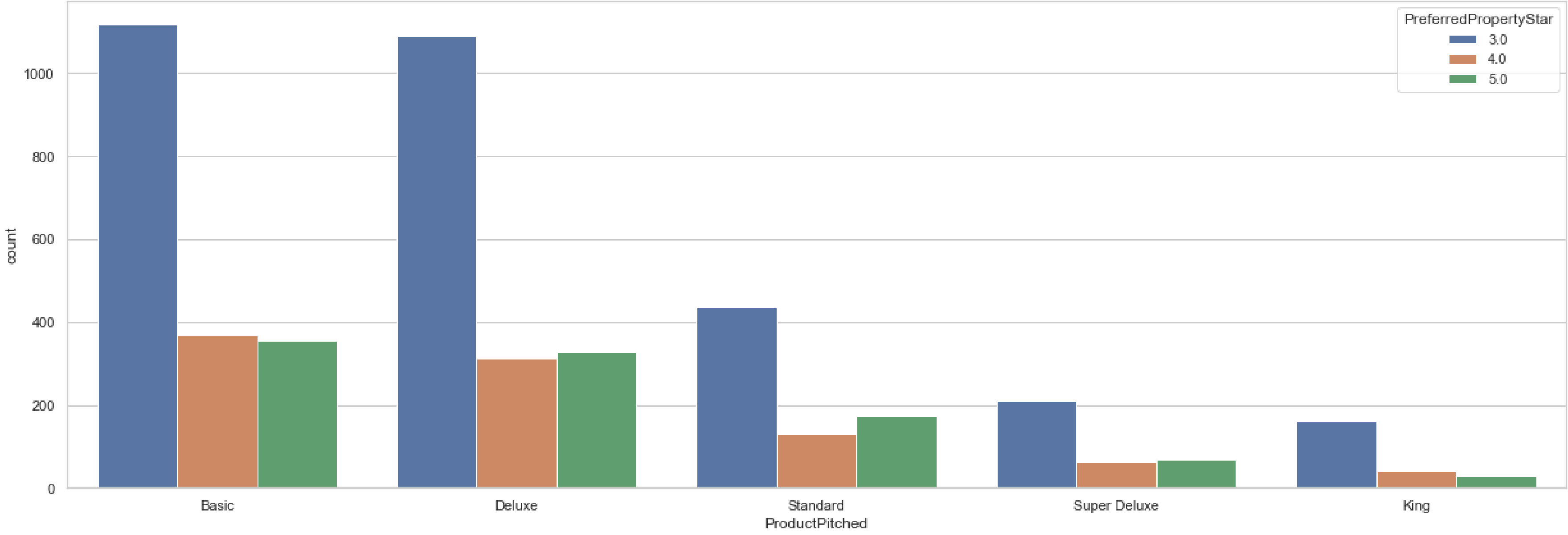
EDA

- The majority of customers were salaried employees with small business in a close second place.

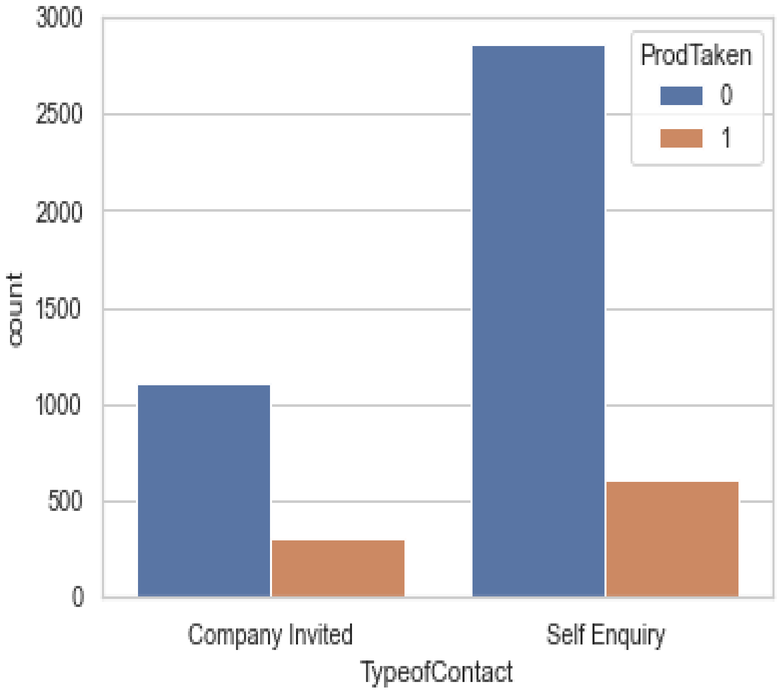


EDA

- The majority of customers rated the Property 3 stars which shows they are only satisfied and not extremely happy with the properties.



EDA

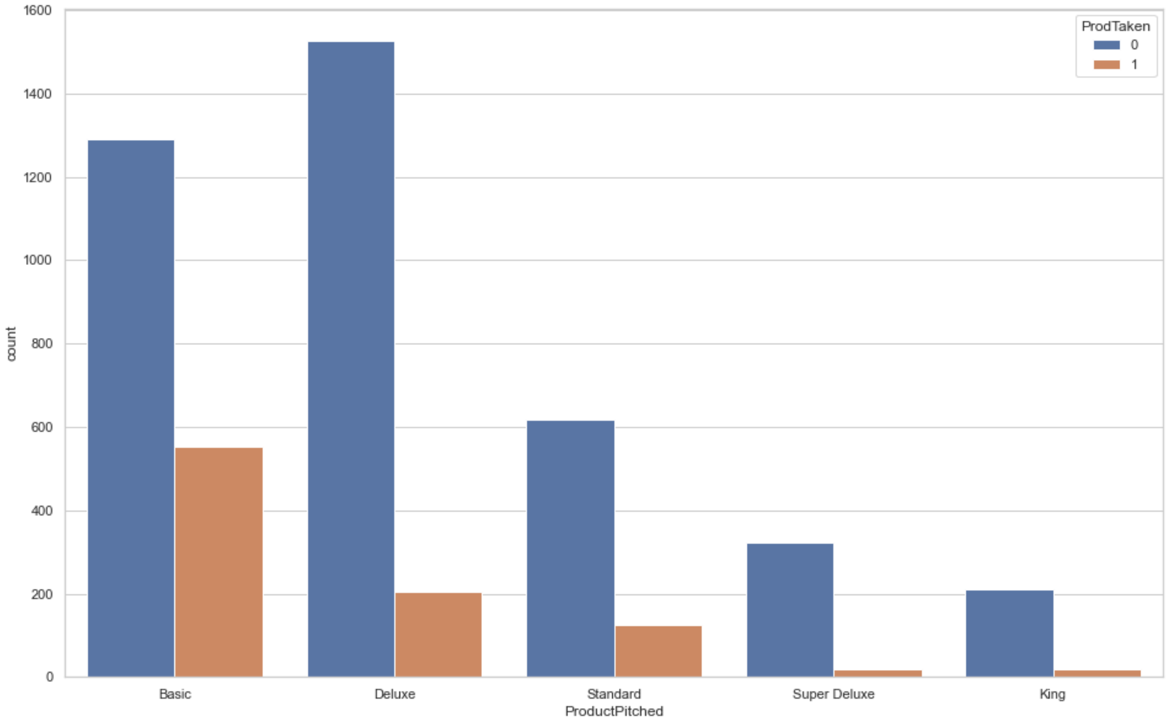


- The turn down rate of the Company Invited is signifcantly much smaller than Self Enquiry.

TypeofContact	ProdTaken	
Company Invited	0	0.226882
	1	0.063421
Self Enquiry	0	0.584902
	1	0.124795

EDA

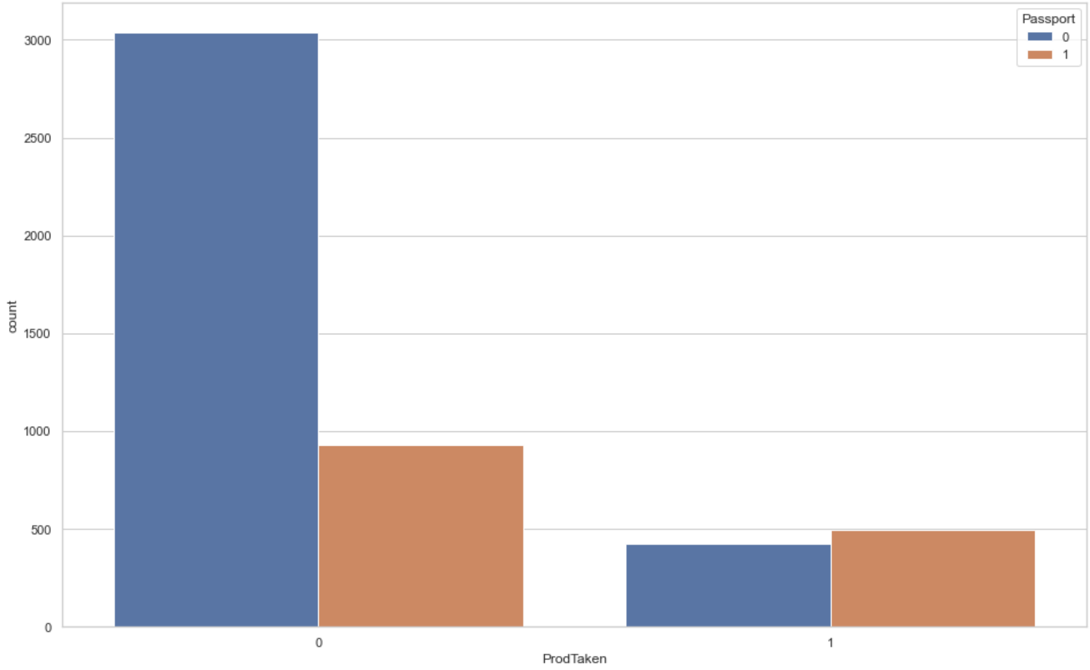
- We can clearly see that the Basic package was accepted the most at a rate of 11% with the next highest rate is only 4%.



ProductPitched	ProdTaken	
Basic	0	0.263912
	1	0.112930
Deluxe	0	0.312602
	1	0.041735
King	0	0.042962
	1	0.004092
Standard	0	0.126432
	1	0.025368
Super Deluxe	0	0.065876
	1	0.004092

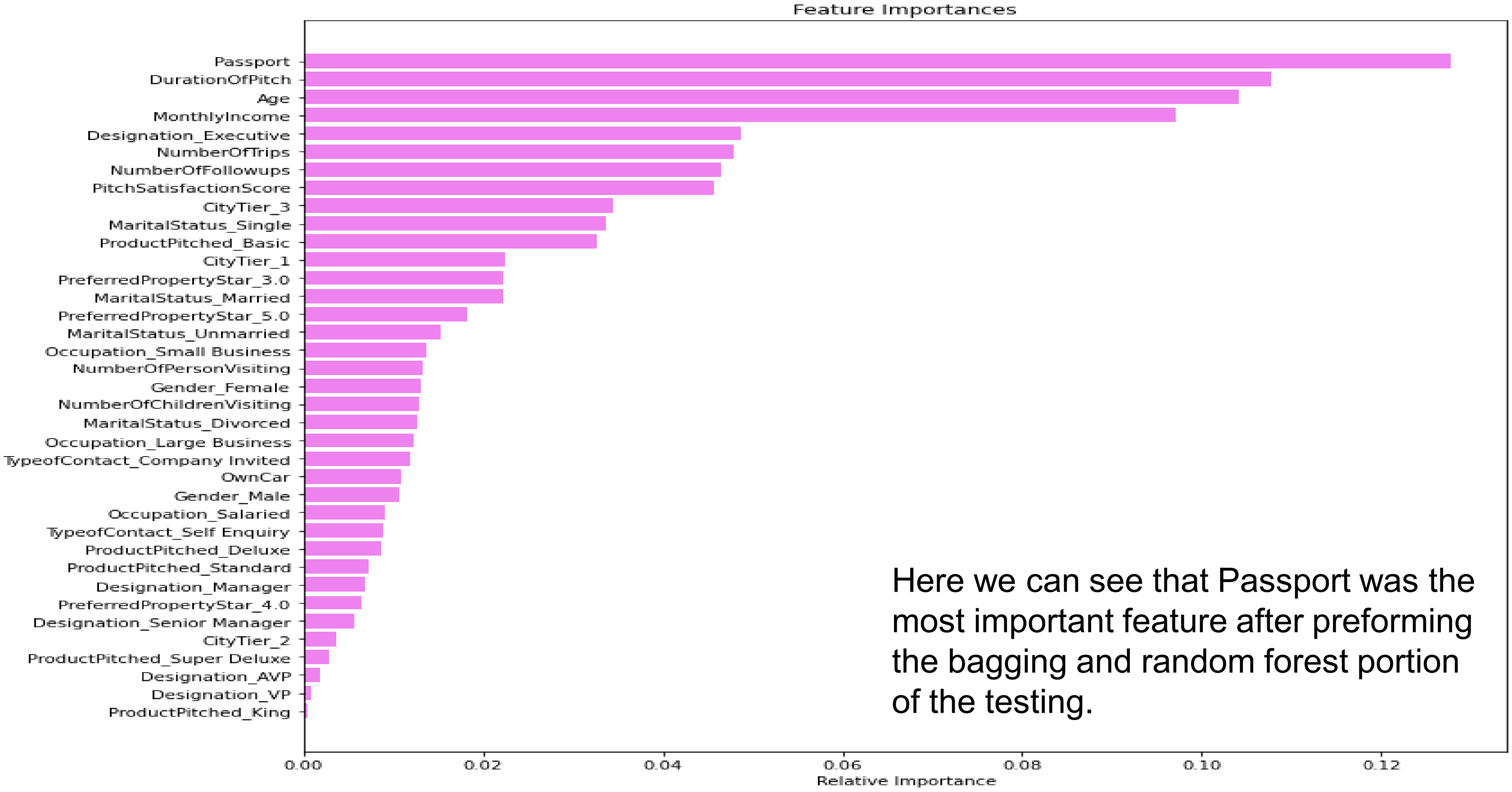
EDA

- This visual shows us that by a small margin people who had a passoport were more likely to buy the product they were offered.



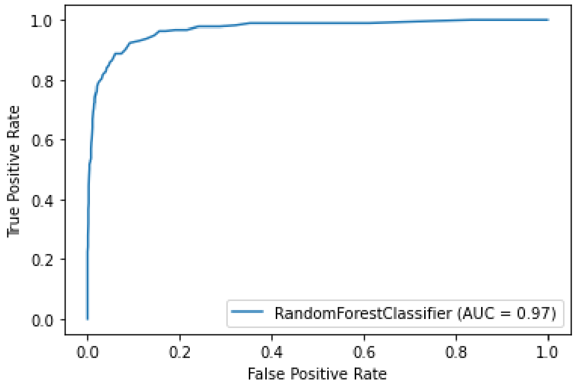
ProdTaken	Passport	
0	0	3040
	1	928
1	0	426
	1	494

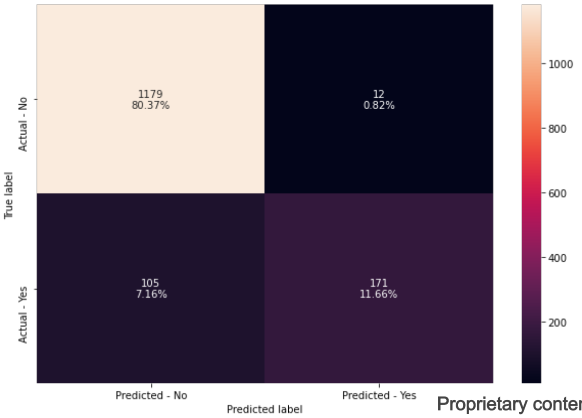
Model Performance Summary On Bagging



Here we can see that Passport was the most important feature after preforming the bagging and random forest portion of the testing.

Model Performance Summary On Bagging





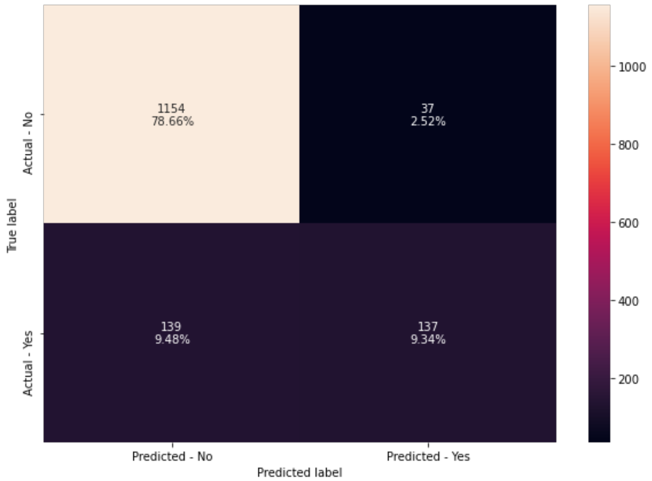
	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision	Train_F1_Score	Test_F1_Score
0	Bagging classifier with default parameters	0.99	0.91	0.97	0.61	1.00	0.88	1.00	0.88
1	Tuned Bagging Classifier	1.00	0.92	1.00	0.63	1.00	0.91	1.00	0.91
2	Bagging classifier with base_estimator=LR	0.81	0.82	0.02	0.03	0.92	0.89	0.92	0.89
3	Random Forest with deafult parameters	1.00	0.92	1.00	0.62	1.00	0.93	1.00	0.93
4	Tuned Random Forest Classifier	0.91	0.88	0.55	0.42	0.95	0.85	0.95	0.85
5	Random Forest with class_weights	0.93	0.87	0.76	0.51	0.85	0.74	0.85	0.74

After comparing all the bagging models, the Random Forest with default parameters had the best performance.

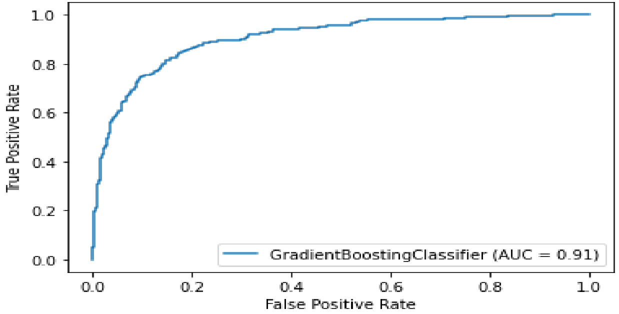
Model Performance Summary on Boosting

- After hypertuning the Adaboost classifier we could see that Monthly Income was the most important feature.
- After hypertuning the Gradient Boosting classifier again Monthly Income was the most important feature.
- After hypertuning the XGBoost classifier Passport was the most important feature.

Model Performance Summary on Boosting



	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision	Train_F1_Score	Test_F1_Score
0	AdaBoost with default paramters	0.84	0.85	0.32	0.33	0.69	0.73	0.69	0.73
1	AdaBoost Tuned	0.97	0.86	0.89	0.61	0.92	0.62	0.92	0.62
2	Gradient Boosting with default parameters	0.89	0.87	0.46	0.41	0.89	0.80	0.89	0.80
3	Gradient Boosting with init=AdaBoost	0.89	0.87	0.46	0.38	0.89	0.81	0.89	0.81
4	Gradient Boosting Tuned	0.92	0.88	0.62	0.50	0.93	0.79	0.93	0.79
5	XGBoost with default parameters	0.88	0.86	0.43	0.36	0.87	0.79	0.87	0.79
6	XGBoost Tuned	0.67	0.67	0.77	0.78	0.33	0.34	0.33	0.34



After comparing all the boosting models, the Gradient Boosting with init=AdaBoost had the best performance.

Business Insights and Recommendations

- With the information that I have and after see that the satisfaction rating with the properties is not great, I would recommend getting rid of the basic package and replace it with the Wellness Package.

- I would also suggest being open to offering any of the products to all of the customers.

- Finally, after seeing the models it is clear that a passport is the most important feature.

- I would recommend that the marketing team target customers the Executive level customers that are married, with at least one child, and a passport.

Business Insights and Recommendations

- Overall, we were able to see that the Executives were the most likely to choose to buy the product, they will more than likely be married, with at least one child.
- We were also able to see that if a customer had a passport, they would be slightly more incline to buy the product.
- I would like more information on why only one product was offered to each Designation. Why could a Deluxe or Standard product be offered to an Executive?
- It is clear that the majority of customers were only satisfied with the property, giving it only a 3-star rating.

greatlearning
Power Ahead

Happy Learning !