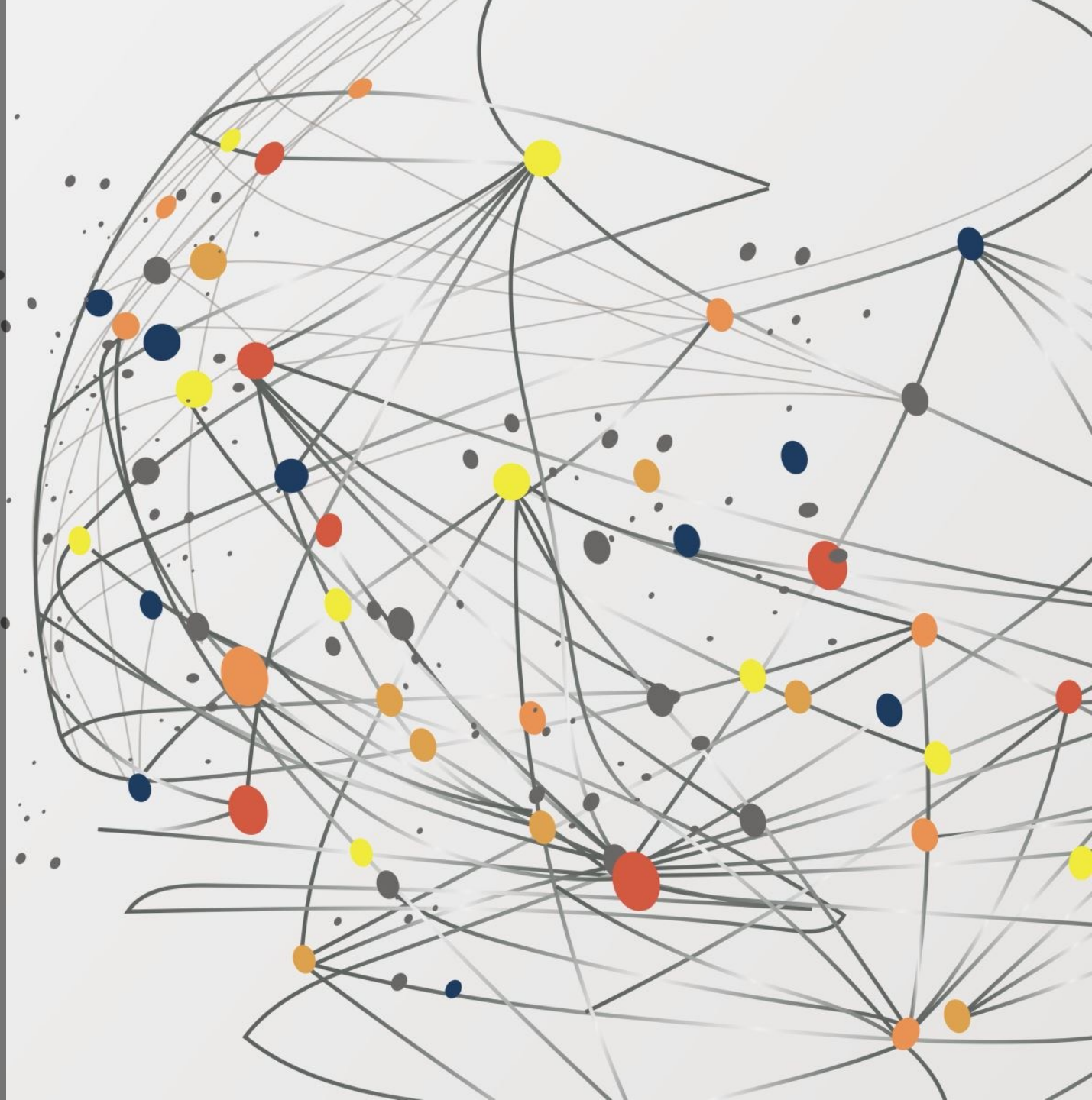
Prediction System – Travel Package Purchase

-Jyothi H



Business Questions

How do we accurately build a travel package purchase prediction system for "Visit with Us"?

Key Questions To Address

The following questions need to be answered to effectively predict if a customer will buy the travel package or not

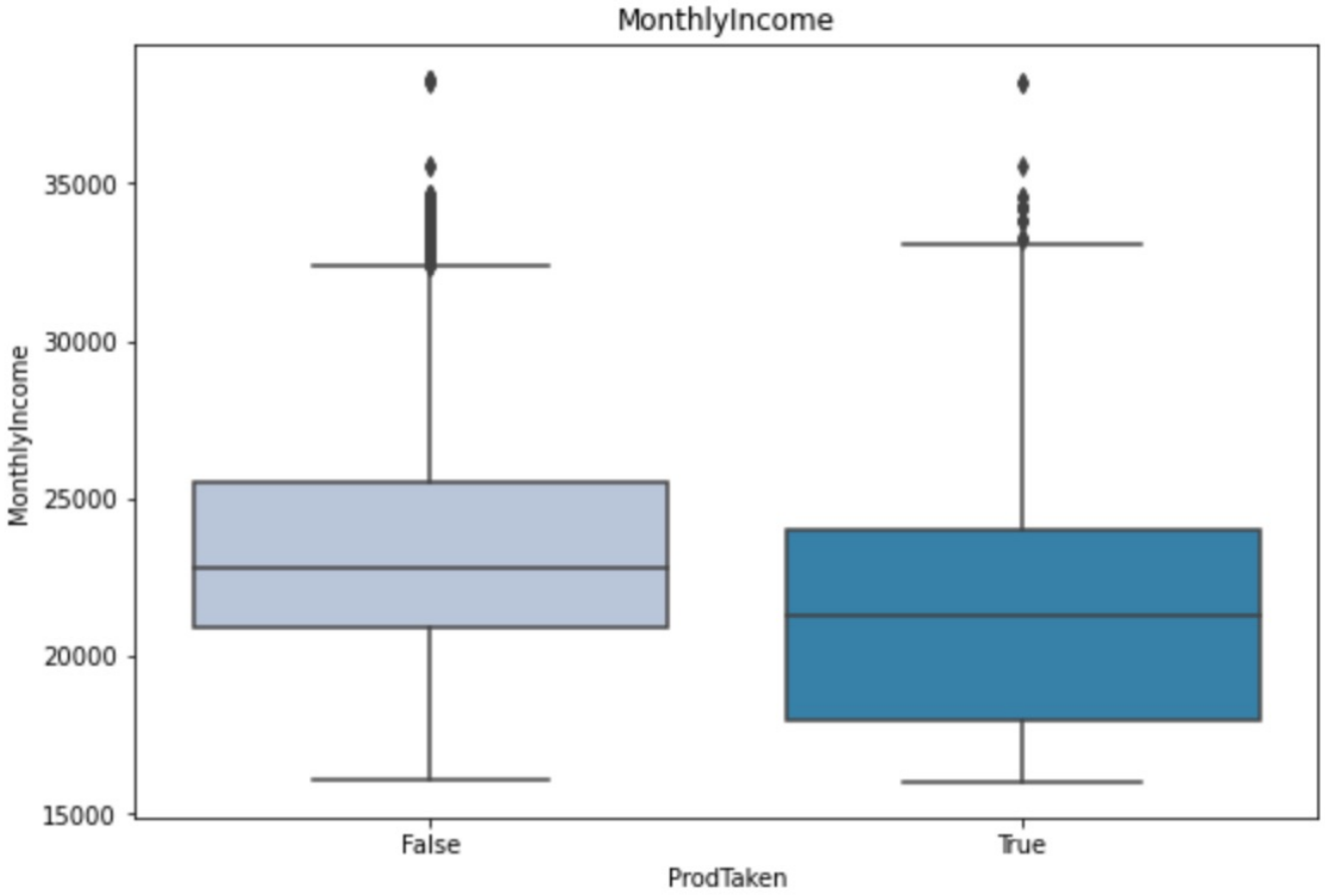
- Does various predicting factors really affect the likelihood of customer purchasing a travel package?
- What are the important variables affecting the likelihood of a customer purchasing a travel package?
- Does the income of the customer really effect the likelihood of purchasing a travel package?
- Does a customer who possesses a passport more likely to purchase the travel package than the one who does not possess a passport?
- Does the number of trips a customer has taken in the past help decide the likelihood of a customer accepting a travel package?
- Does the number of person visiting really effect the likelihood of purchasing a travel package?
- Does the duration of sales pitch effect the likelihood of purchasing a travel package?
- Do designation really effect the likelihood of customer purchasing a travel package?

Classification Techniques & Models Used For Prediction

- Bagging Classification Models
  - Decision Tree Estimator
  - Random Forest Estimator
  - Bagging Estimator with Tuning
  - Random Forest with Tuning
  - Logistic Regression with Bagging Classifier
  - Random Forest with Weighted Score

- Boosting Classification Models
  - AdaBoost Classifier
  - GradientBoost Classifier
  - XGBoost Classifier
  - AdaBoost Classifier with Tuning
  - GradientBoost Classifier with Tuning

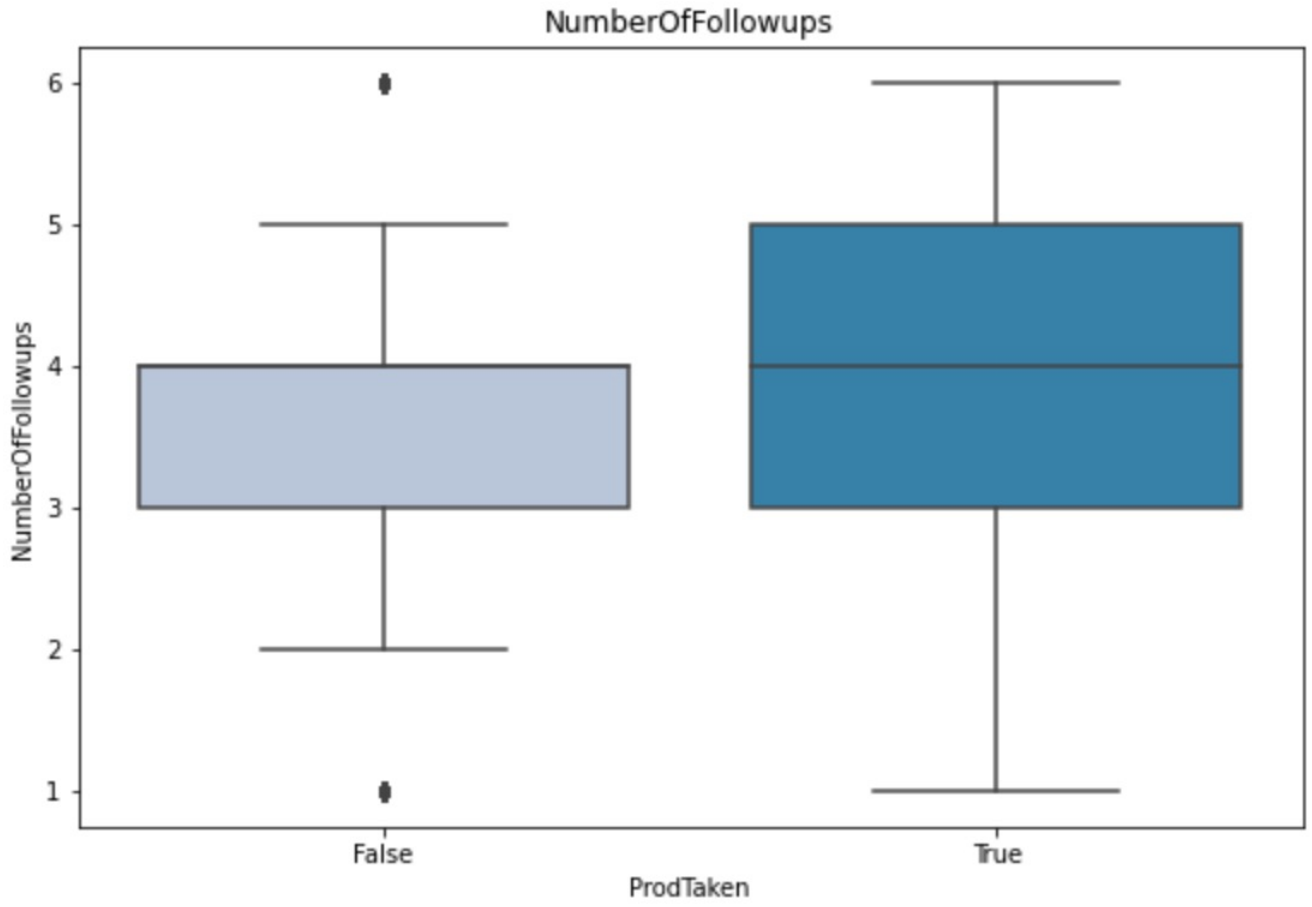
Correlation – Product Purchase Vs Income



As depicted in the graph, the likelihood of a customer buying a travel package tends to be positively correlated with the Income of a customer.

Therefore, we would expect to see this as an important feature in our model prediction.

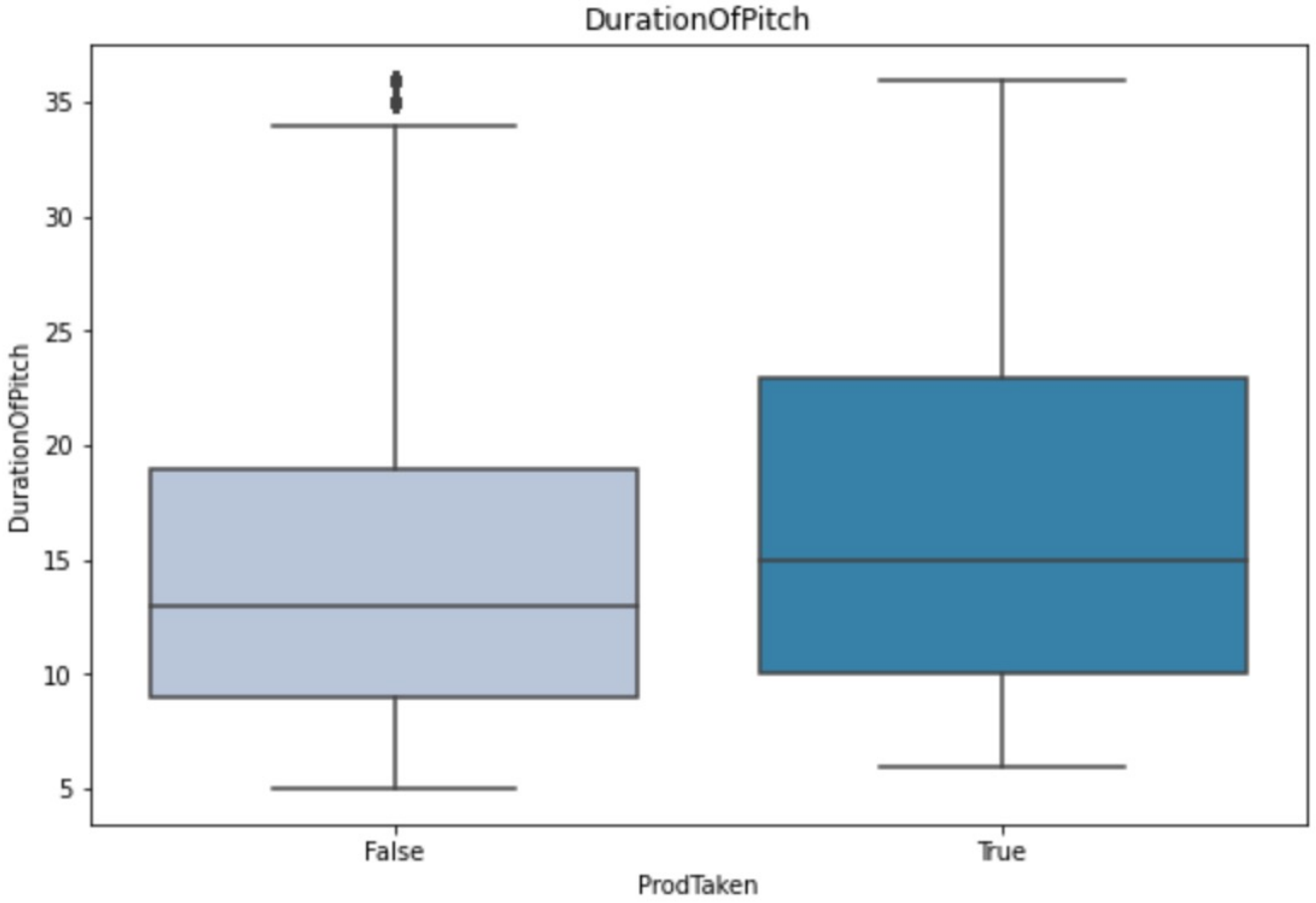
Correlation – Product Purchase Vs Number of Follow-ups



The graph shows that a higher number of follow-ups is associated with a higher likelihood of travel package purchase by the customer.

Therefore, we should expect to see number of follow-ups as an important feature in our model prediction.

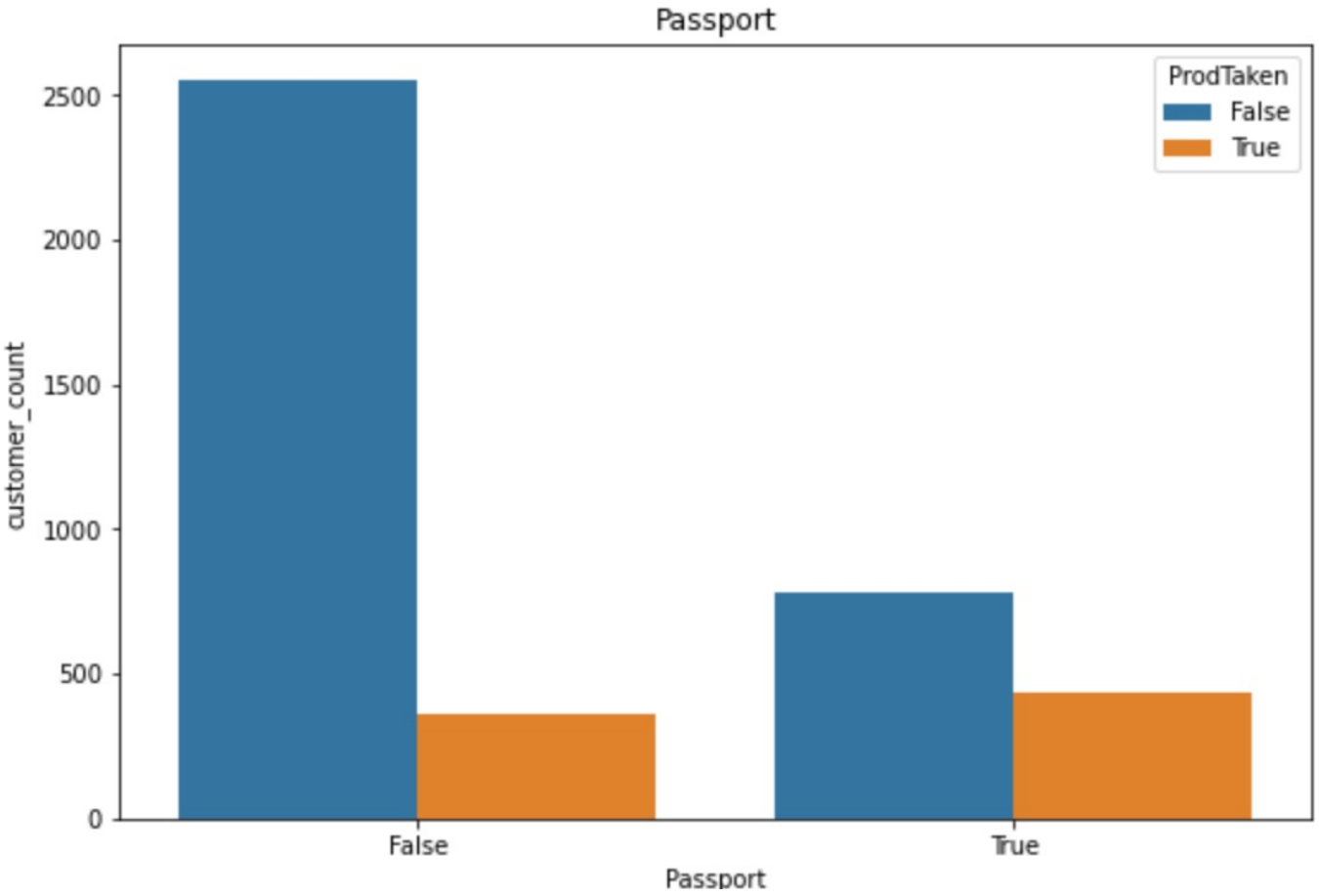
Correlation – Product Purchase Vs Duration of Sales Pitch



The graph shows that a duration of sales pitch is associated with a higher likelihood of travel purchase by the customer.

Therefore, we should expect to see duration of sales pitch as an important feature in our model prediction.

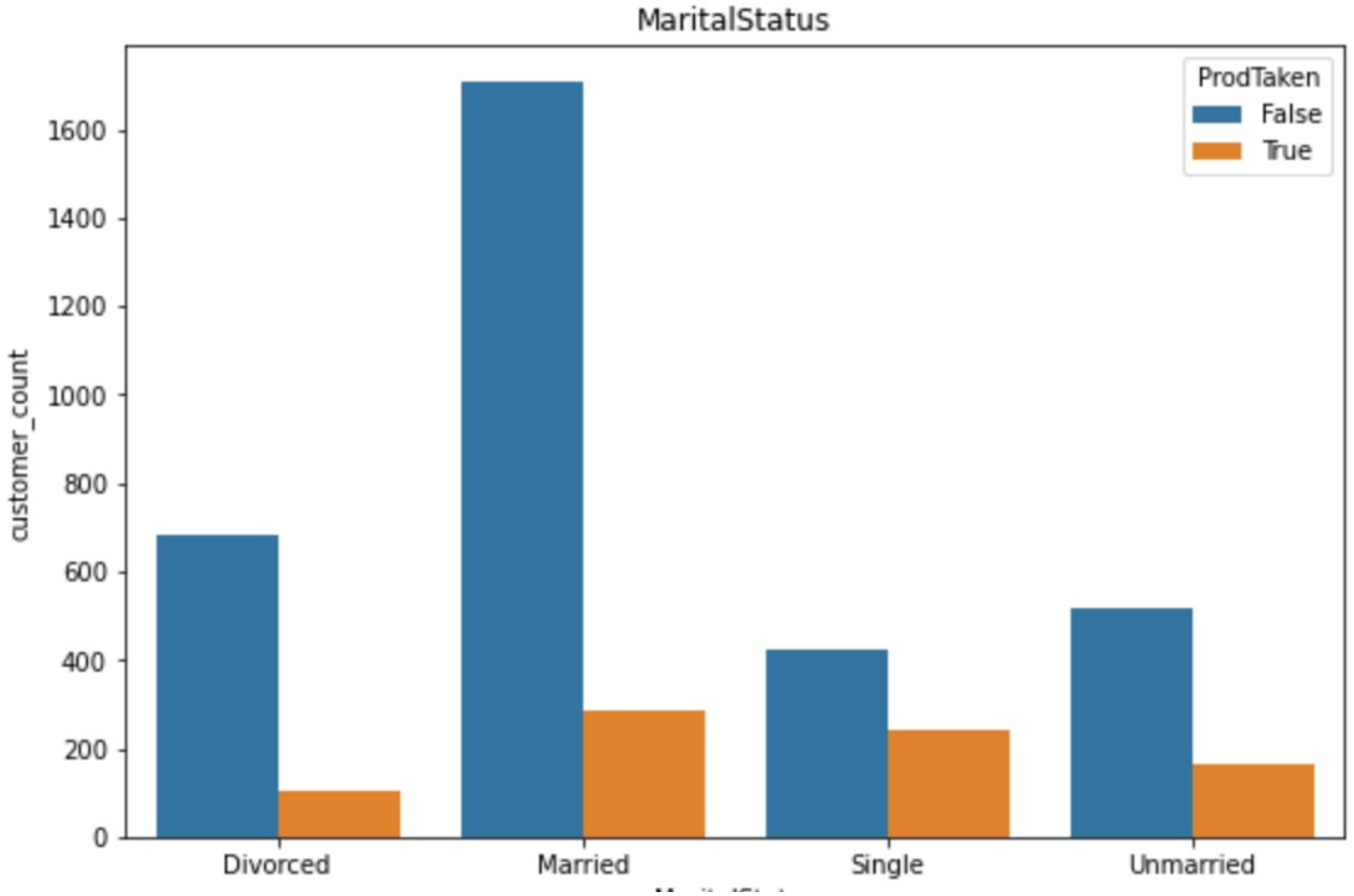
Correlation – Product Purchase Vs “Posses Passport?”



20.0% of the users who did not accept travel package have passport and 50.0% of the users who accepted travel package have passport. This implies that customers who own a passport are more likely to accept the travel package

Therefore, we should expect to see passport attribute as an important feature in our model prediction.

Correlation – Product Purchase Vs Marital Status



70.0% of the users who did not accept travel package are either married or divorced & 50.0% of the users who accepted travel package are either married or divorced. This implies that married and divorced people are less likely to accept travel package than single or unmarried people

Therefore, we should expect to see passport attribute as an important feature in our model prediction.

Correlation (EDA) – Categorical Attributes

Features	Did Not Accept Past Product/Travel Package	Did Accept Past Product/Travel Package	Significance
Has Passport	20%	50%	It appears that Possessing a passport does seem to increase the likelihood of a customer will purchase the travel package
Gender = Male	60%	60%	60% of the customers who did not accept the travel package were Male and 60% of the customers who accepted the travel package were Male, implying that Gender does not seem to have bearing on travel package acceptance
Marital Status = Married/Divorced ?	70%	50%	70% of customer who rejected the travel package were married/divorced and 50% of the customers who accepted the package were married or divorced. Marital status seems to have a bearing on the likelihood of customer accepting a travel package
Own Car?	60%	60%	Car ownership status does not seem to have a bearing on the customer loan acceptance

Model Evaluation Criteria

Possible Errors in Prediction:

1. Error 1 - Predicting a customer will buy a travel package but in reality, does not.
    - Leads to misuse of marketing expenditure and resources
2. Error 2 - Predicting a customer will not buy a travel package loan but in reality, he will –
    - Results in potential lost opportunity

Which Case Holds More Weight ?

- Error 1 results in misuse of capital which can be critical for small cash strapped companies
- Error 2 results in lost opportunity, which is important for stable companies with healthy balance sheet, looking to grow.

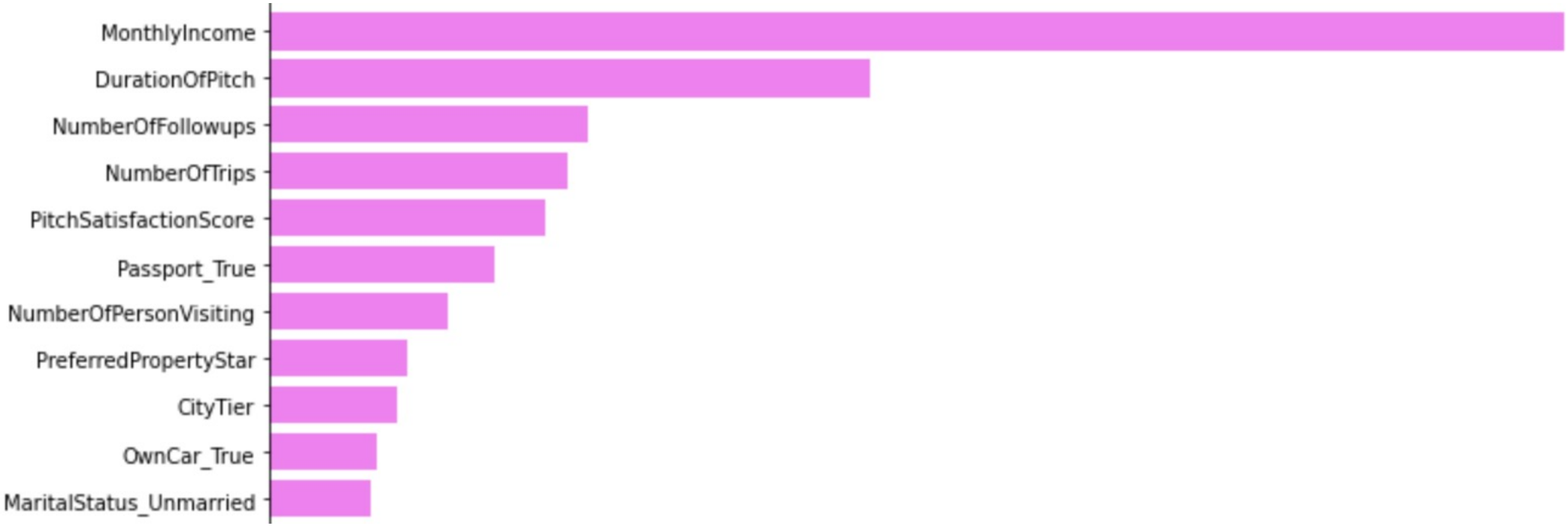
We will assume Error 1 to be important and therefore will try to maximize the Recall Score to get an optimal model

Model Prediction

	Model	Accuracy	Recall	Precision	F1 Score
Training Data	Decision Tree Estimator	0.993763	0.967742	1	0.983607
	Random Forest Tuned	0.940055	0.777778	0.898551	0.833814
	Random Forest With Weighted Score	0.910949	0.770609	0.769231	0.769919
	XGBoost Classifier	0.999653	0.998208	1	0.999103
	AdaBoost Classifier Tuned	0.978170	0.926523	0.959184	0.942571
Testing Data	Decision Tree Estimator	0.894184	0.598326	0.803371	0.685851
	Random Forest Tuned	0.859451	0.548117	0.664975	0.600917
	Random Forest With Weighted Score	0.84895	0.552301	0.622642	0.585366
	XGBoost Classifier	0.905493	0.65272	0.821053	0.727273
	AdaBoost Classifier Tuned	0.877221	0.610879	0.712195	0.657658

Best Model Recommendation : AdaBoost Classifier Tuned, since it gives a better balance of a generalized model and performance

Model Prediction – Relative Feature Importance



Potential Benefits Of Accurate Prediction Model

- Increased Marketing ROI – By accurately targeting the right segment of customers, the marketing team can maximize its ROI on marketing spend.

- Increased Conversion Rate – Accurate prediction model allows the company to target the right customers , thereby increasing the operational KPI's of the company , mainly the conversion wrt product sale.

- Improved Profit Margin – Optimal use of resource and capital that is targeted towards intended customers helps company in improving the profit margin.

Recommendation

The Boosting model with AdaBoost Classifier with tuning results in a superior model with a good balance of generalization and model performance, given a set of features that are described by the regression model. Therefore, I recommend leveraging this model using the below mentioned key factors to determine the likelihood of customer accepting a travel package

- Customer Income
- Duration of Sales Pitch
- Number of Follow-ups with Customer
- Historical Yearly Trips Taken by Customer
- Customer Passport Possession Status
- Marital Status of Customer

In conclusion, I recommend the marketing team to target a customer segment with higher income, who possess a passport, with higher historical trips taken by customer , and who is single/unmarried for its campaigning. Also, an increase in the duration of sales pitch along with follow-ups , increase the likelihood of customer purchasing the travel product