VISIT WITH US – PROJECT 5

Sarah A. Thomas

BUSINESS PROBLEM OVERVIEW AND SOLUTION APPROACH

- Currently, there are 5 travel packages offered through Visit With Us: Basic, Standard, Deluxe, Super Deluxe, and King. Last year, 18% of potential customers purchased a package.
- A new package is being planned called the Wellness Tourism Package that allows the traveler to maintain, enhance, or kick-start a healthy lifestyle. Is this a viable package? What should the marketing strategy be?
- The goal of the Data Science team is to generate a statistical model using Python to figure out which data points can help predict who of the potential customer base is likely to purchase a package. The team will use a variety of Bagging and Boosting techniques to construct the models.

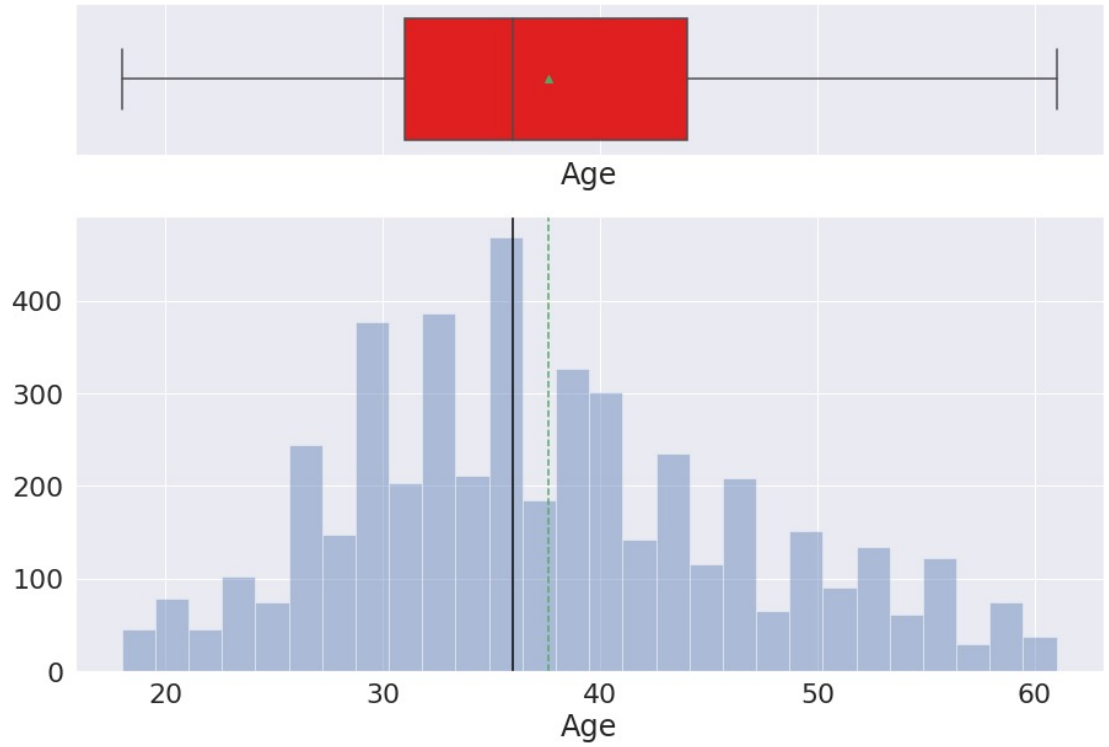
DATA OVERVIEW (1 OF 2)

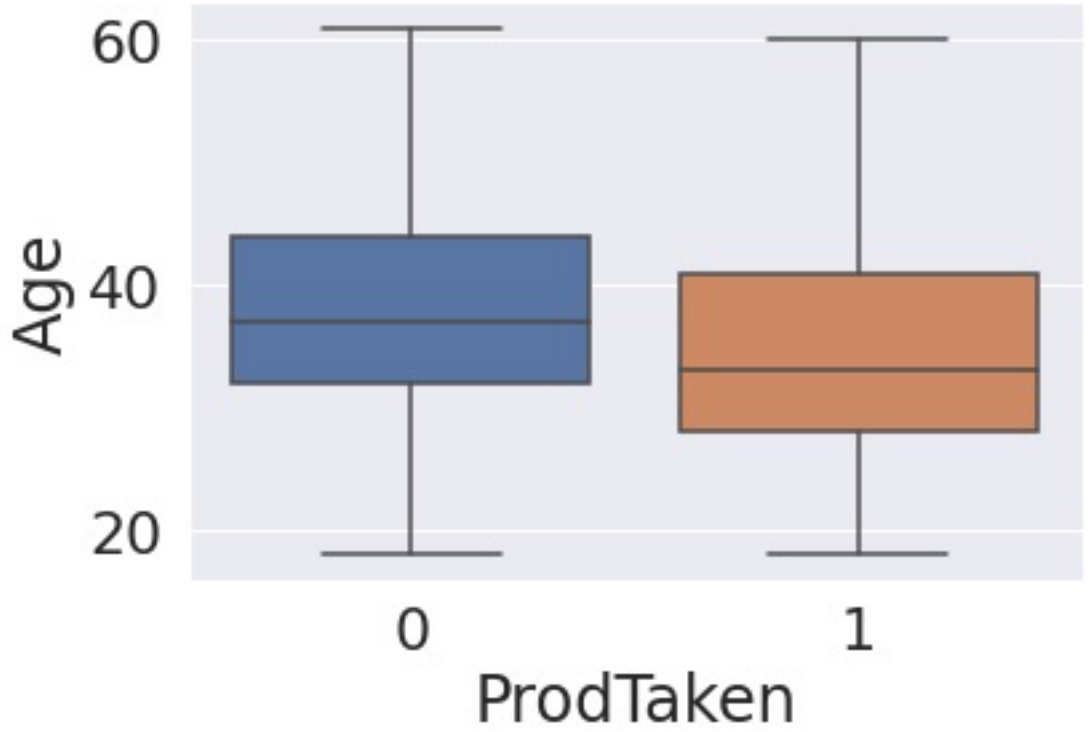
Variable	Description
CustomerID	Unique Customer ID
ProdTaken	Whether the customer has purchased a package (0:No, 1:Yes)
Age	Age of customer
TypeofContact	How customer was contacted (Company Invited or Self Enquiry)
CityTier	Depends on the development of a city, population, facilities, and living standards (Tier 1 > Tier 2 > Tier 3)
Occupation	Occupation of customer
Gender	Gender of customer
NumberOfPersonVisiting	Total number of persons planning to take trip with the customer
PreferredPropertyStar	Preferred hotel property rating by customer
MaritalStatus	Marital status of customer

DATA OVERVIEW (2 OF 2)

Variable	Description
NumberOfTrips	Average number of trips in a year by customer
Passport	Whether the customer has a passport (0:No, 1:Yes)
OwnCar	Whether the customer owns a car (0:No, 1:Yes)
NumberOfChildrenVisiting	Number of children <5 yrs planning to take trip with customer
Designation	Designation of the customer in the current organization
MonthlyIncome	Gross monthly income of the customer
PitchSatisfactionScore	Sales pitch satisfaction score
ProductPitched	Product pitched by the salesperson
NumberOfFollowups	Total number of follow-ups done by the salesperson after the pitch
DurationOfPitch	Duration of sales pitch

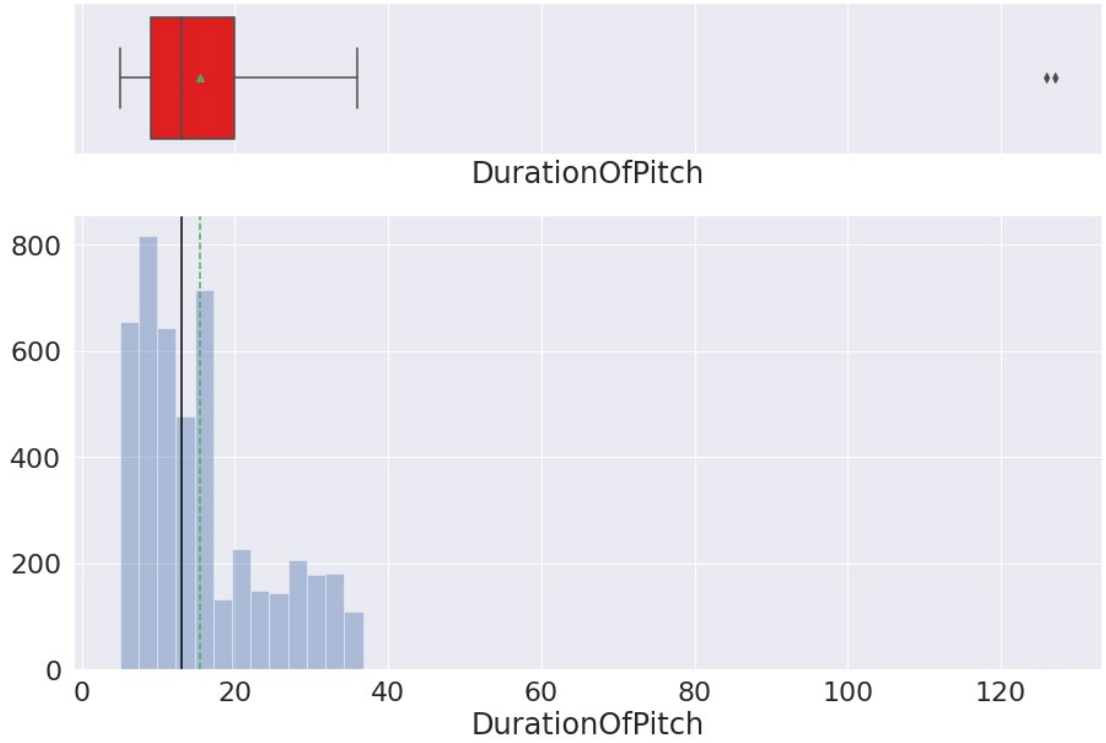
AGE

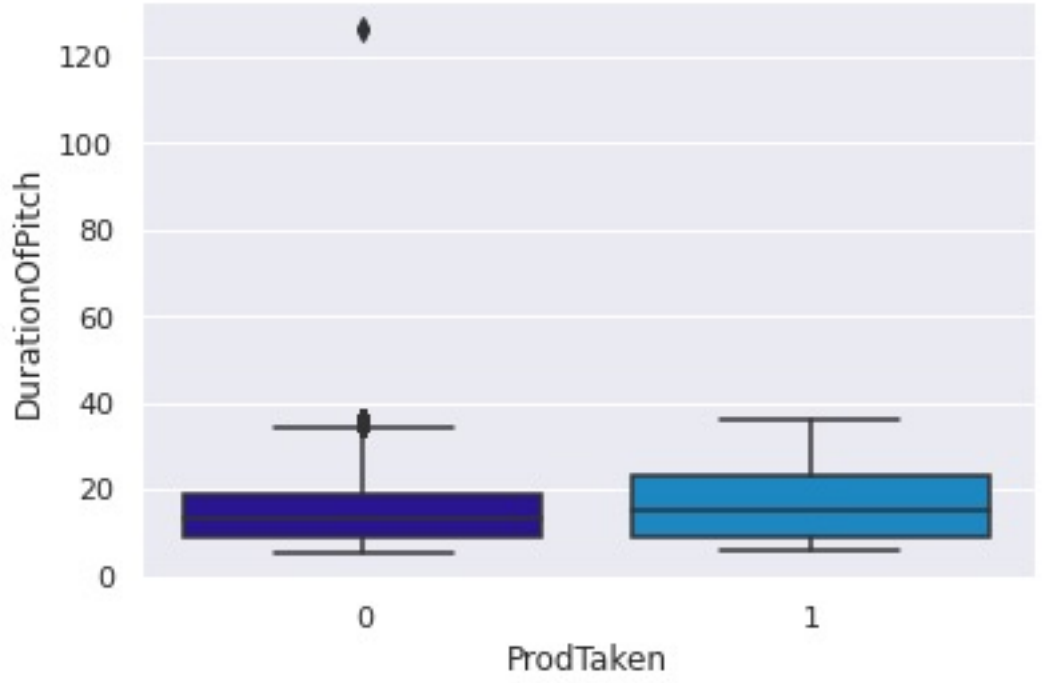




Age ranges from 18 to 61 with a median age of 36 and mean of 37.62. Indicates right-skewed distribution. Customers who purchase travel package tend to be slightly younger.

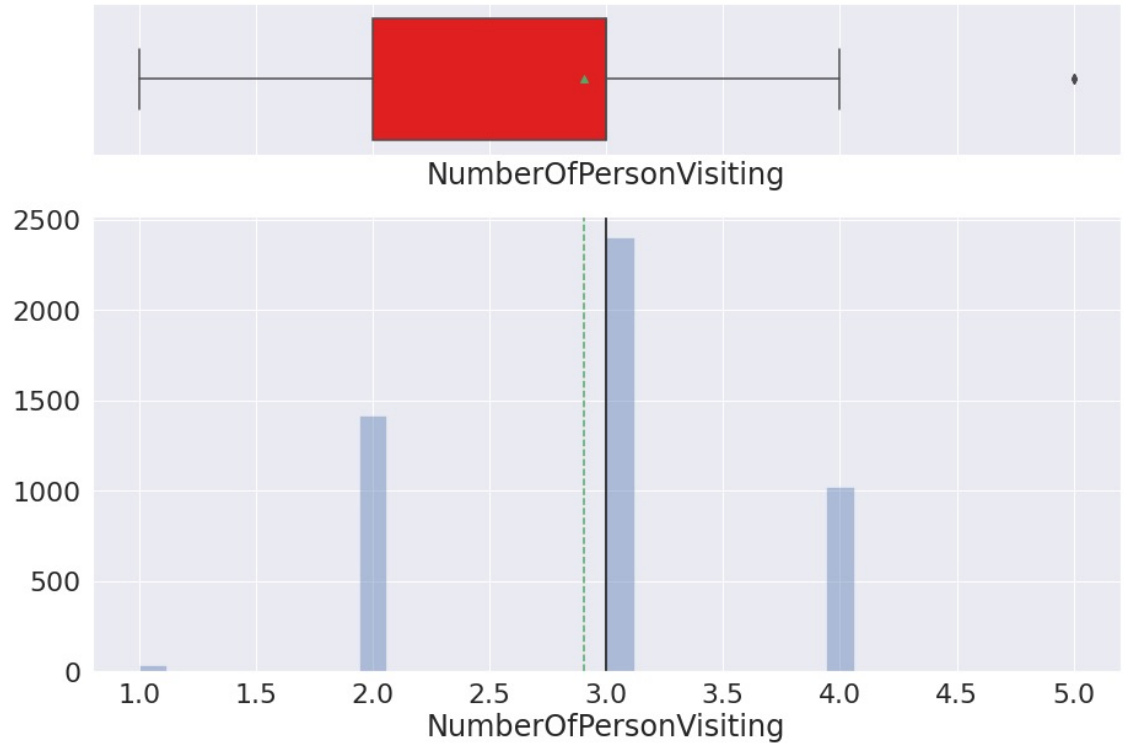
DURATION OF PITCH

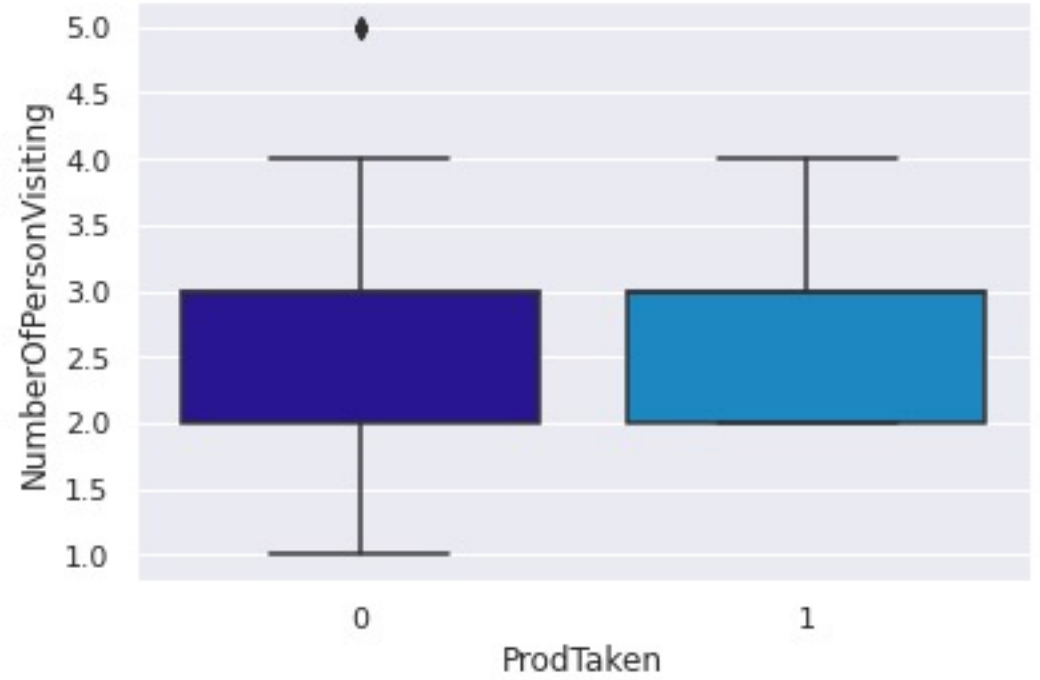




The mean is 15.49 minutes and the median is 13. Indicates a right-skewed distribution. Much longer pitches did not result in packages being purchased.

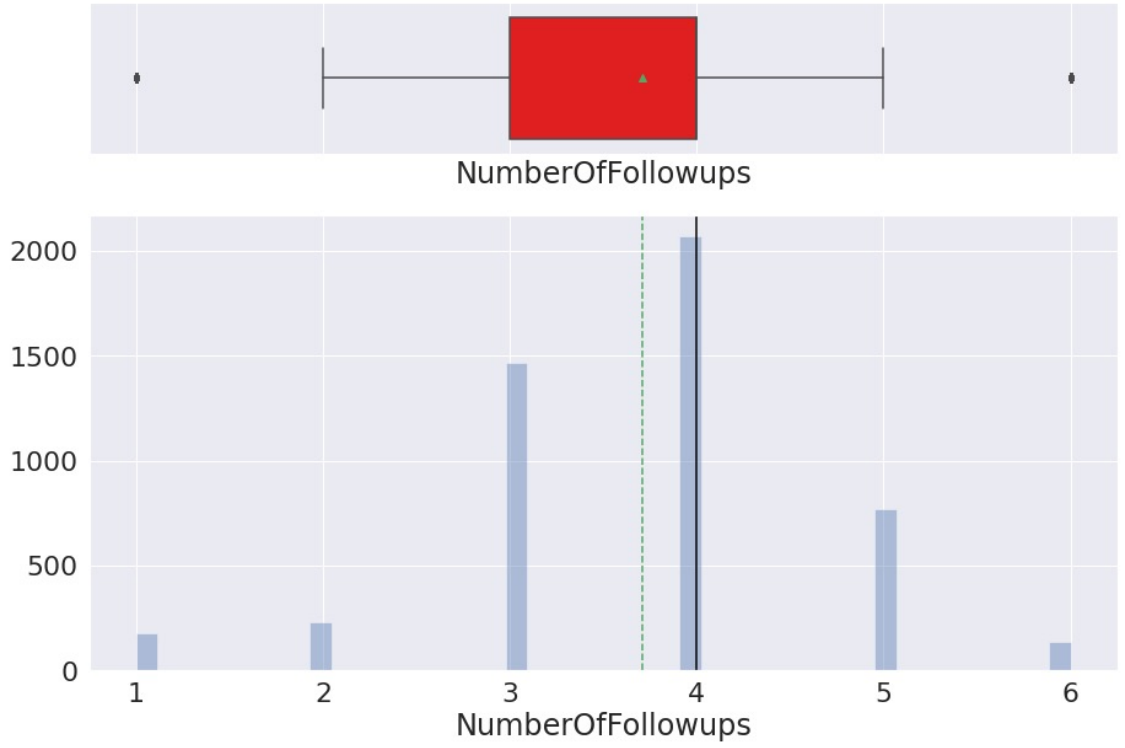
NUMBER OF PERSONS VISITING

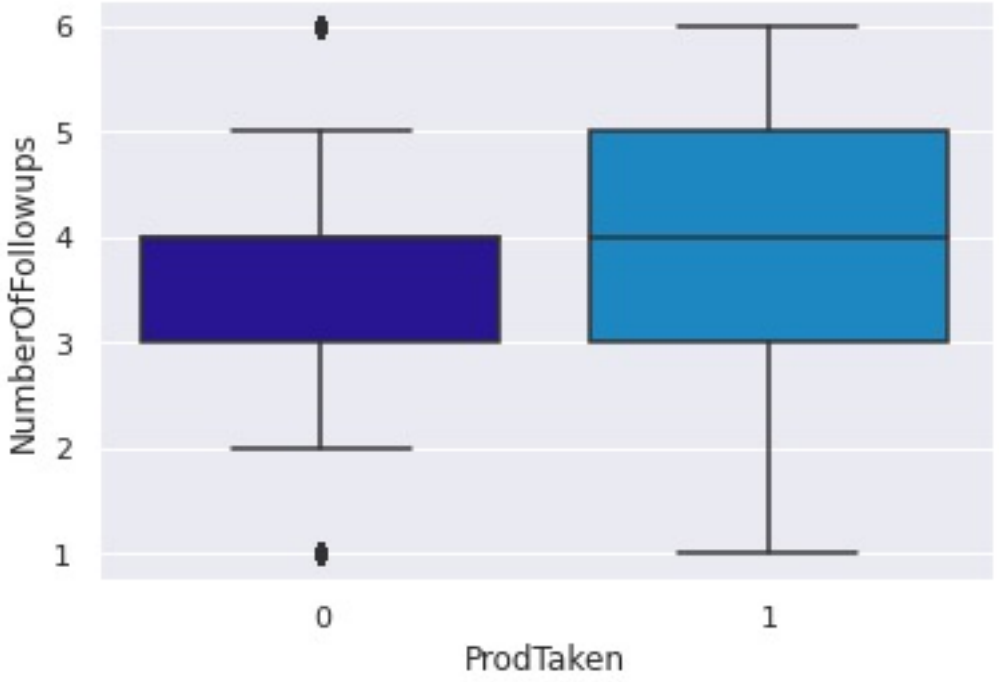




The mean for number of persons planning to take a trip with the customer is 2.91 with a median of 3. Indicates a nearly symmetrical distribution. The number of persons going on the trip does not seem to make a difference in packages purchased.

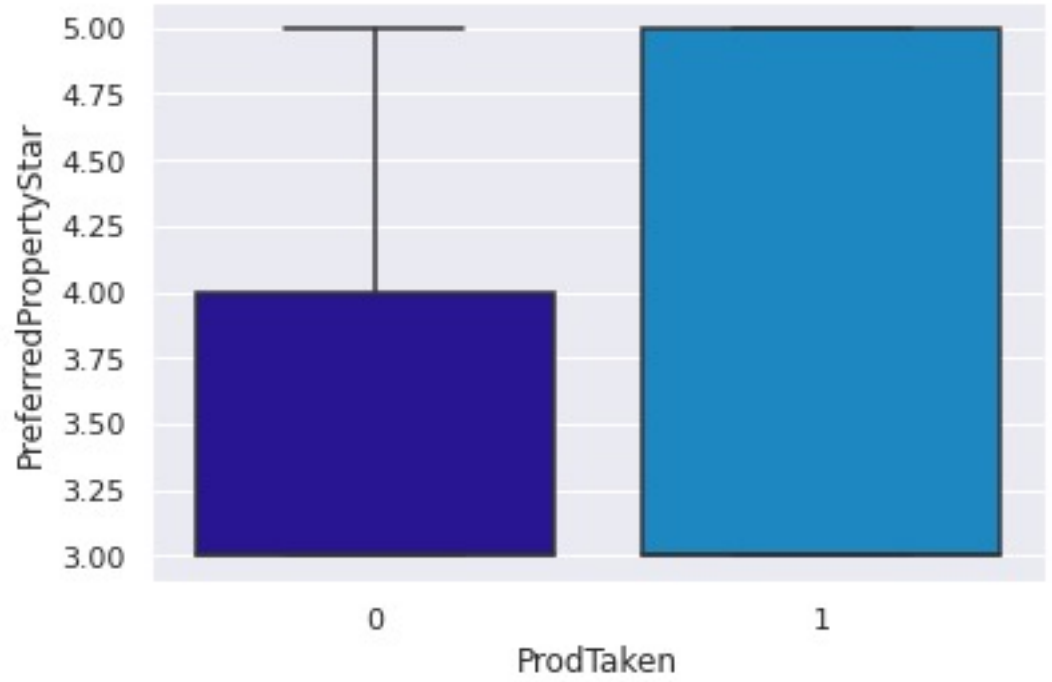
NUMBER OF FOLLOW-UPS





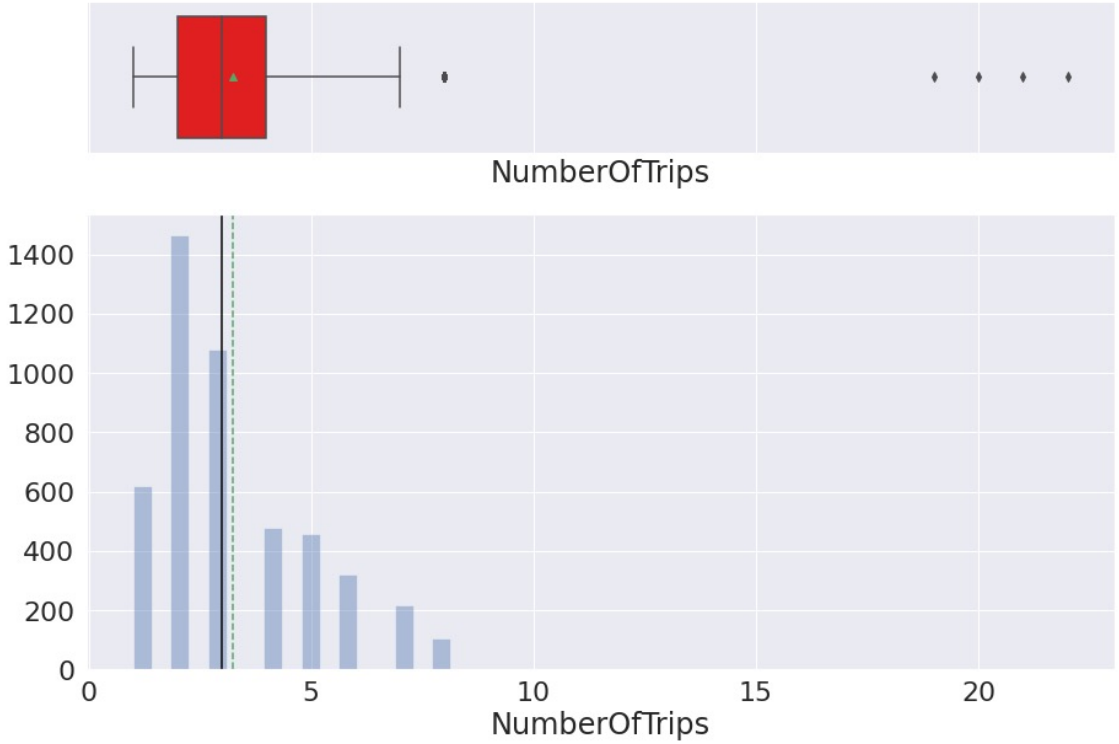
The mean number of follow-ups after the sales pitch is 3.71 and the median is 4.0, indicating a nearly symmetrical distribution. The number of followups appear to make a positive difference in the number of packages sold.

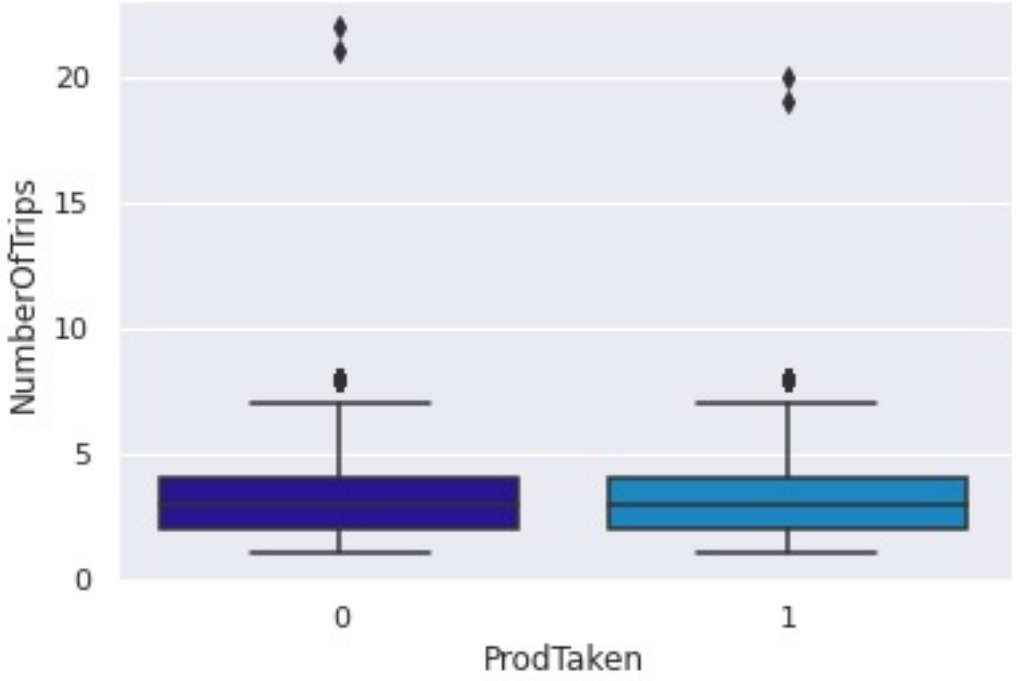
PREFERRED PROPERTY STAR



The mean customer rating was 3.58 and the median was 3.0, indicating a slightly right-skewed distribution. Those who purchased packages tended to rate properties across a wider range (from lower to higher) in terms of property stars.

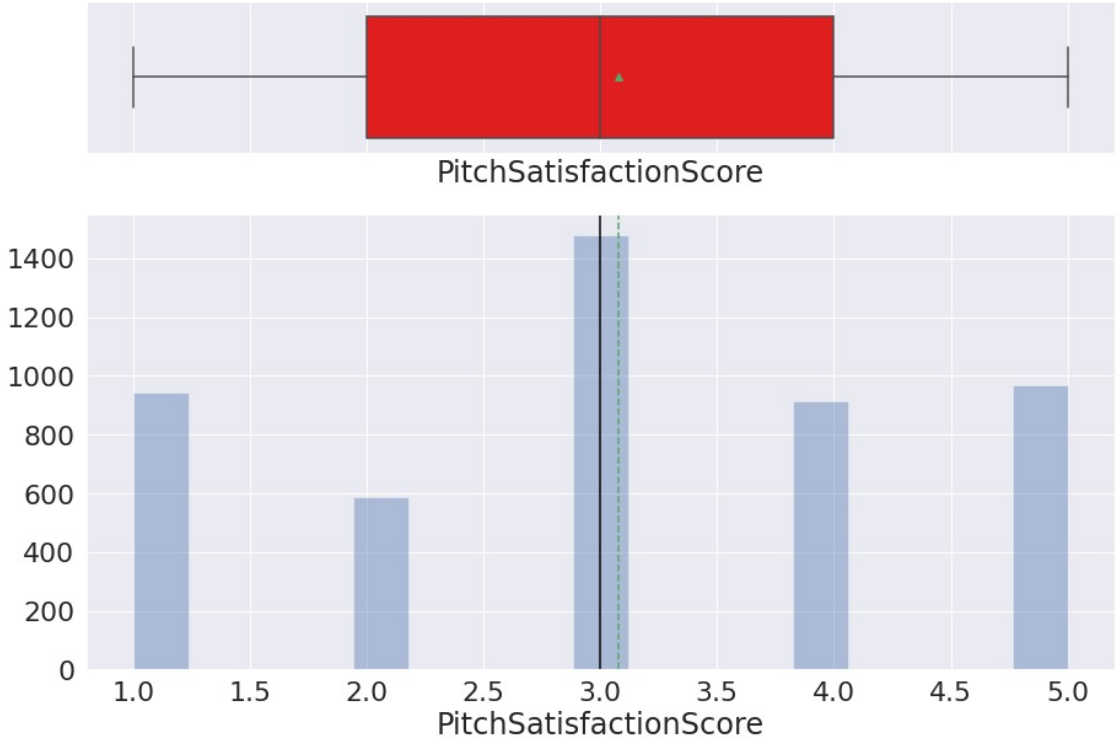
NUMBER OF TRIPS

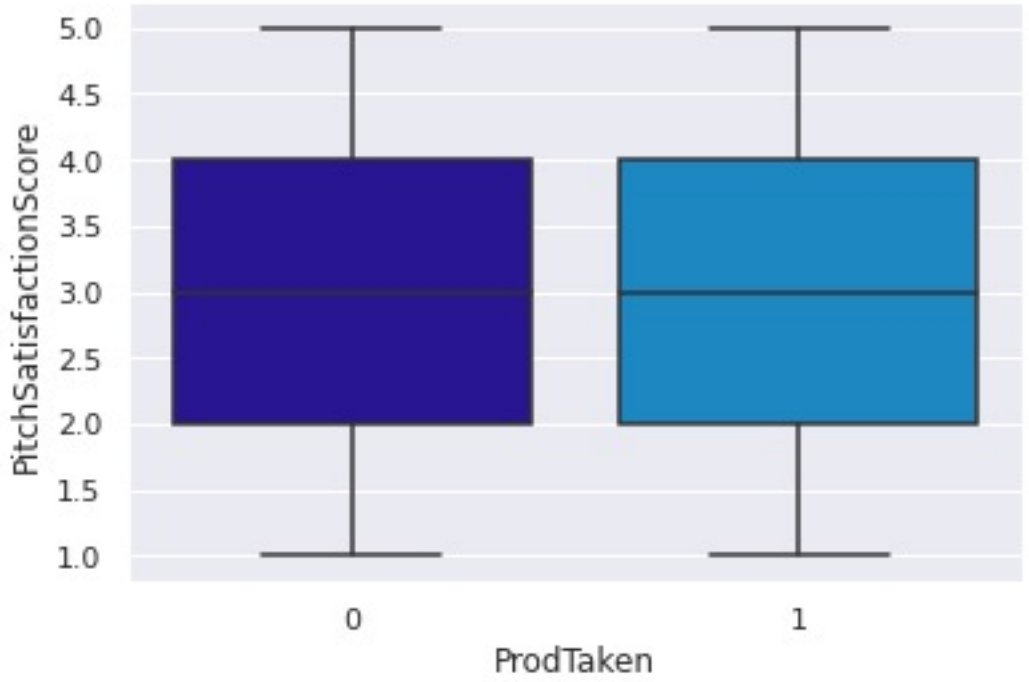




The mean is 3.24 and the median is 3 indicating a nearly symmetrical distribution. NumberOfTrips does not seem to have a significant impact on whether packages are purchased.

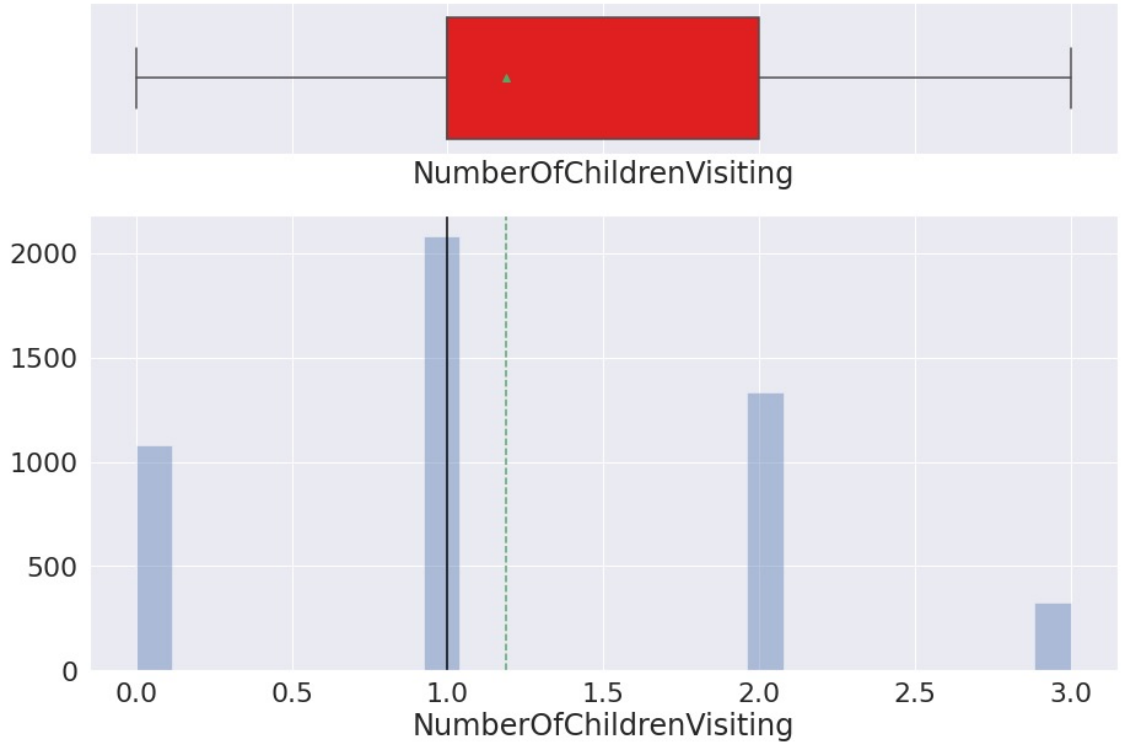
PITCH SATISFACTION SCORE

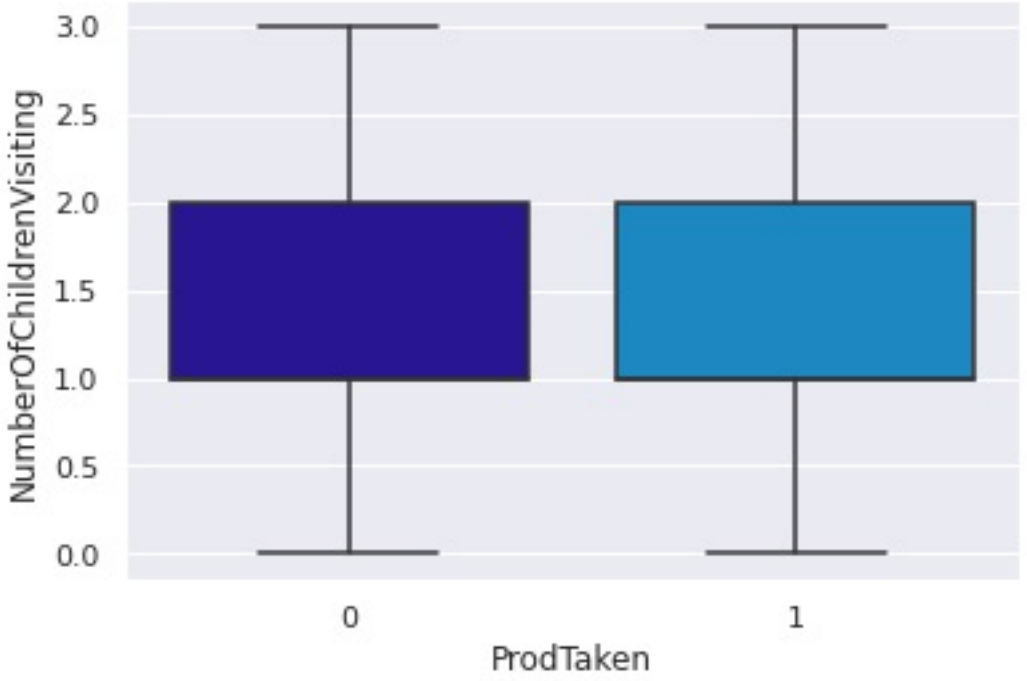




The mean is 3.08 and the median is 3.0, indicating a nearly symmetrical distribution. PitchSatisfactionScore doesn't appear to have a great impact on whether or not a package is purchased.

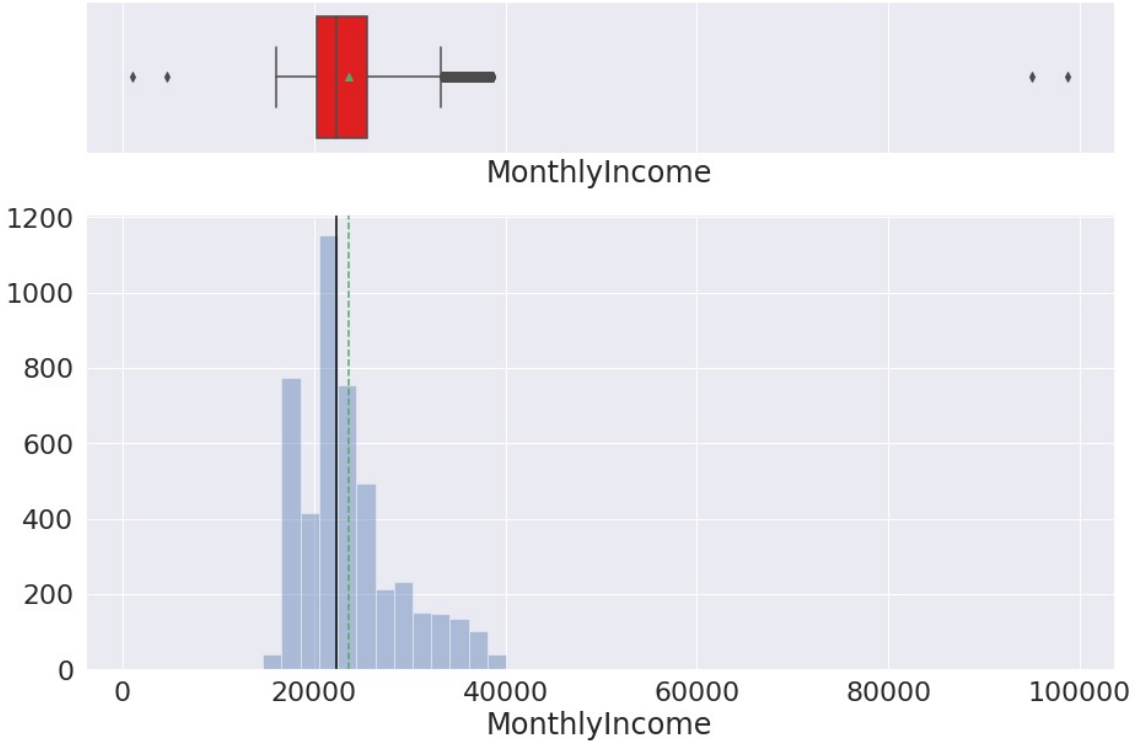
NUMBER OF CHILDREN VISITING

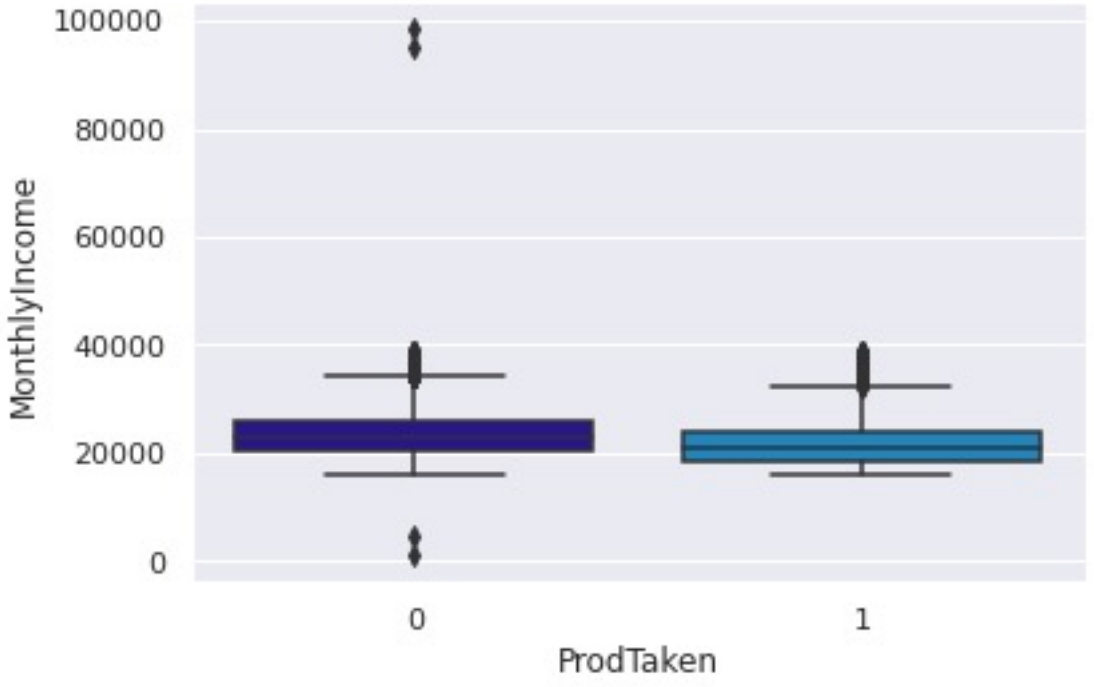




Number of children less than age 5 planning to take the trip with the customer - mean is 1.19 and median is 1.0 indicating a nearly symmetrical distribution. NumberOfChildrenVisiting does not appear to have an impact on whether or not a customer will purchase a package.

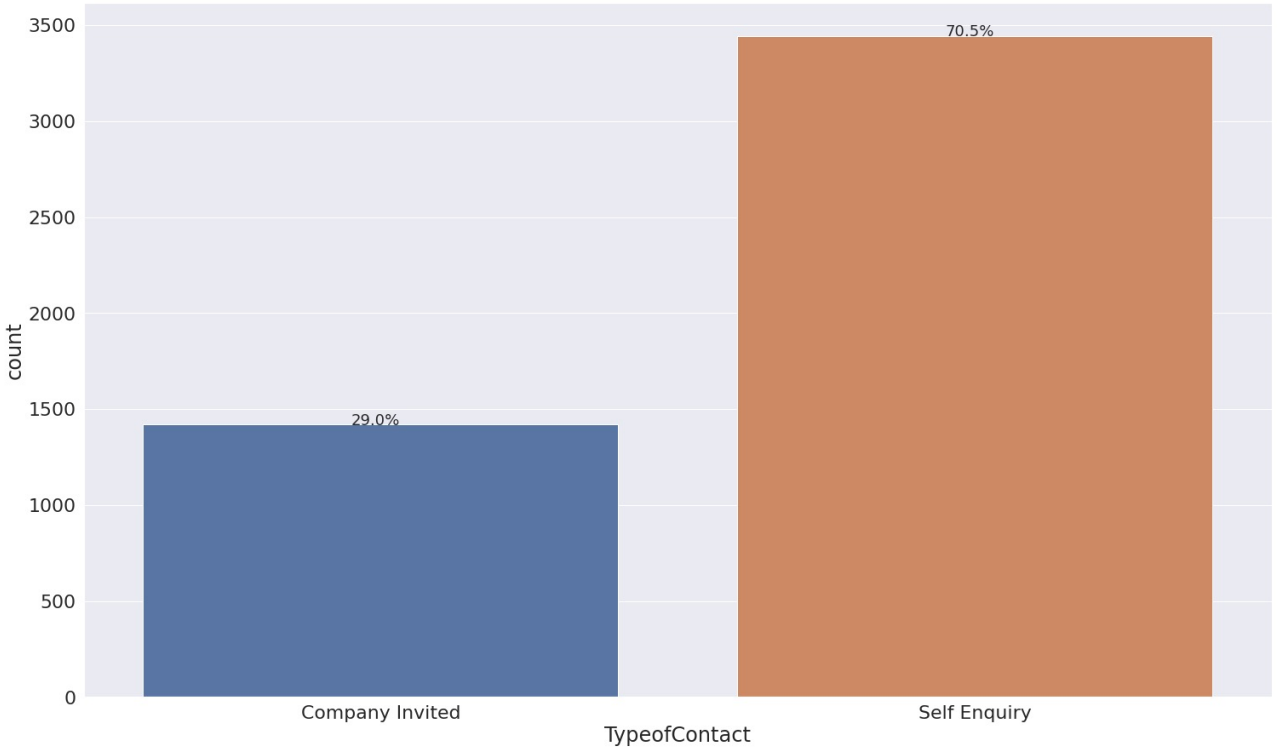
MONTHLY INCOME

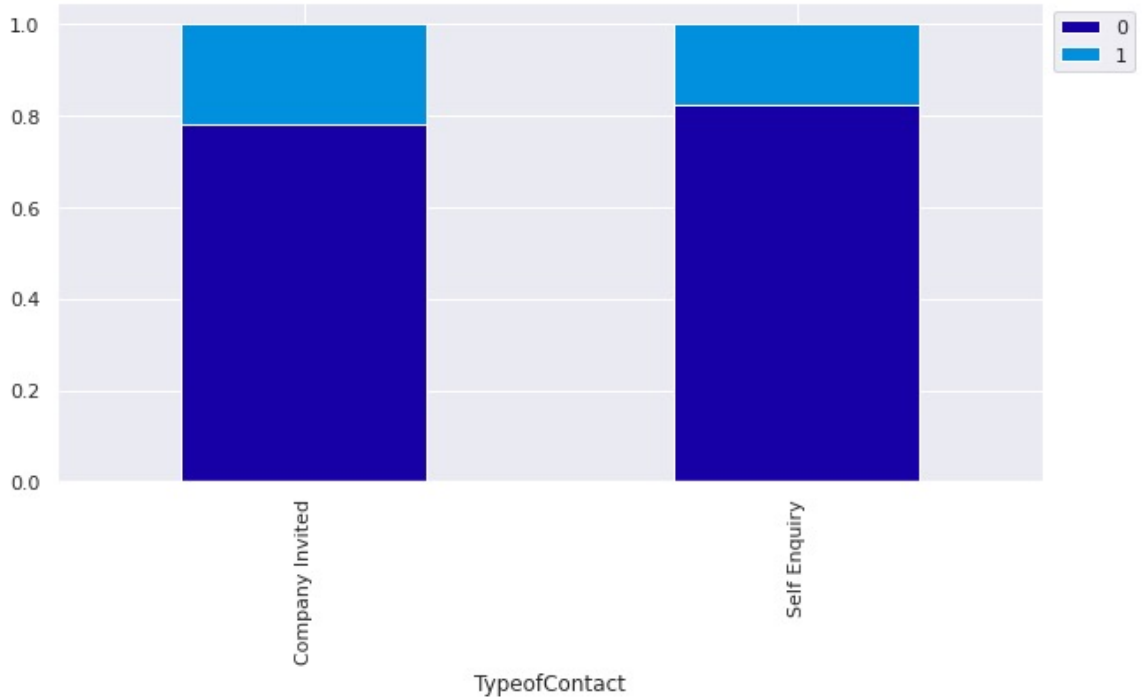




Mean is 23619.85 and median is 22347 indicating a right-skewed distribution. A much higher income (looking in paricular at the outliers) does not dictate whether a package will be purchased.

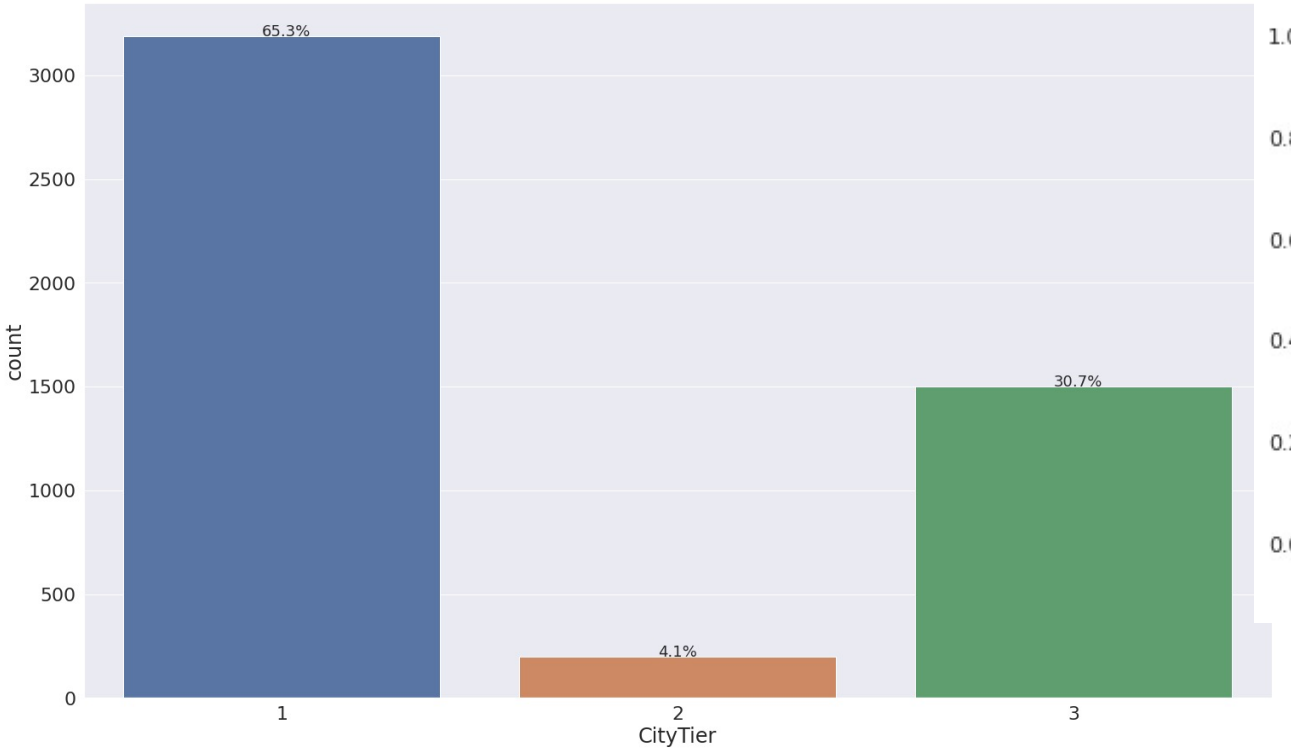
TYPE OF CONTACT

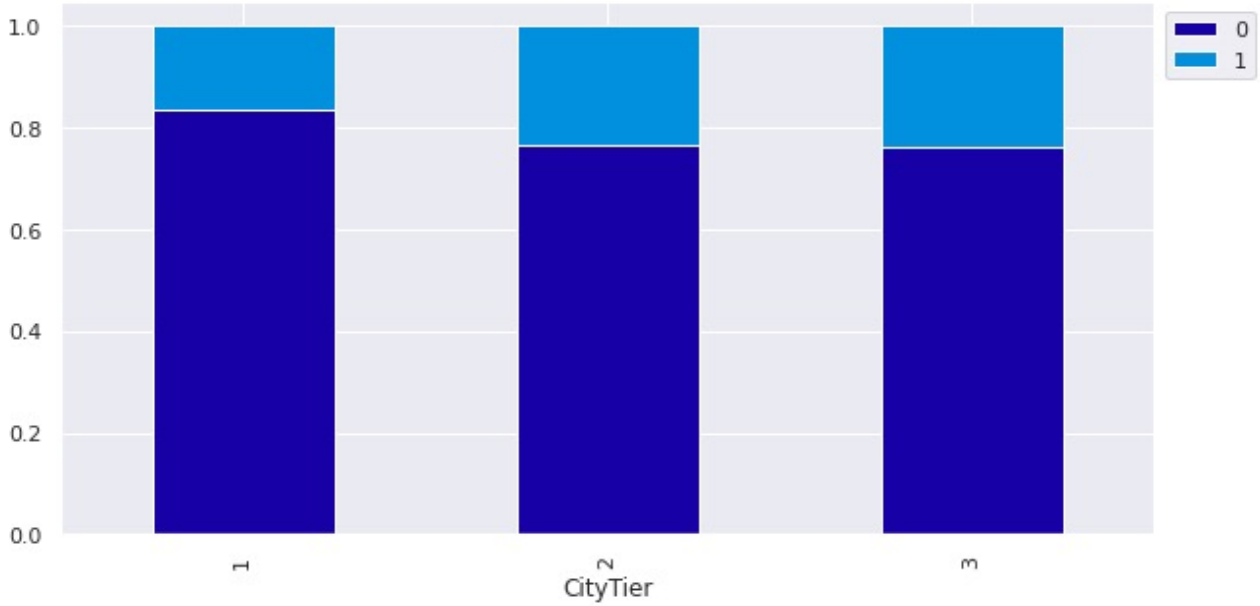




Most customers were self enquiries. There does not seem to be a major difference in the number of packages sold depending on the type of contact (company invited or self inquiry).

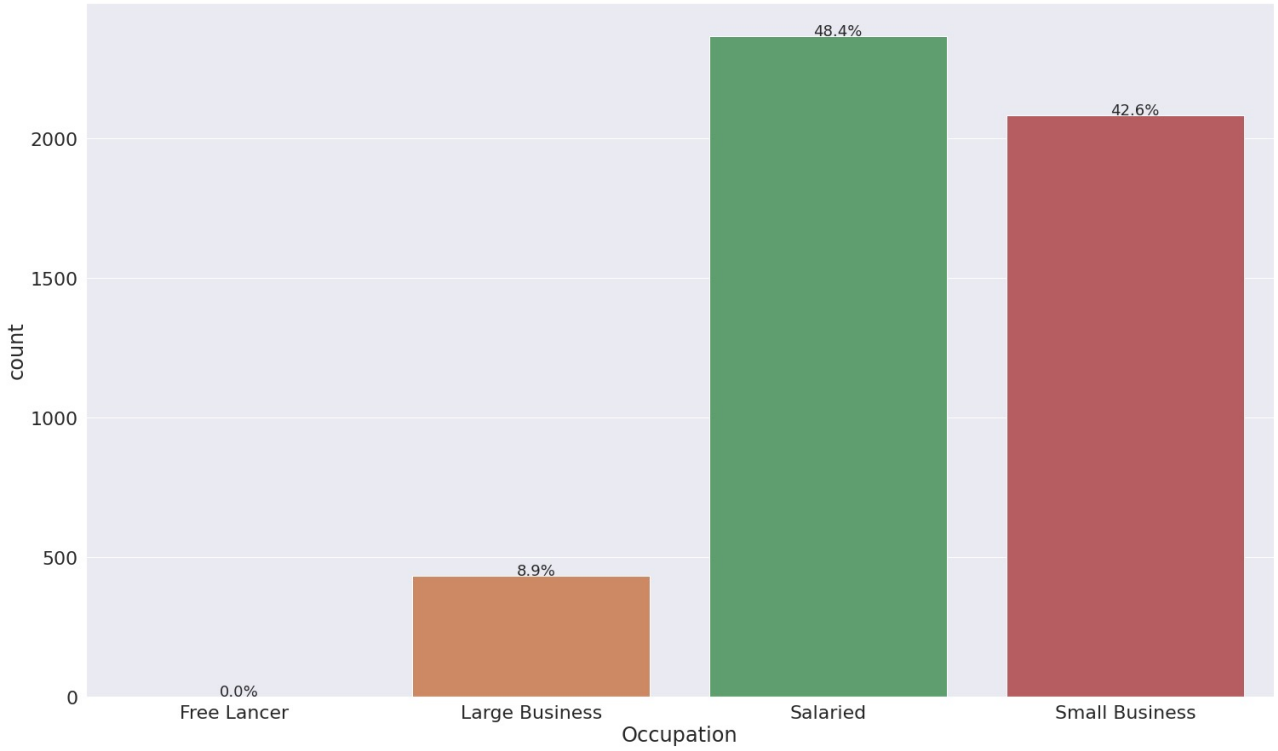
CITY TIER

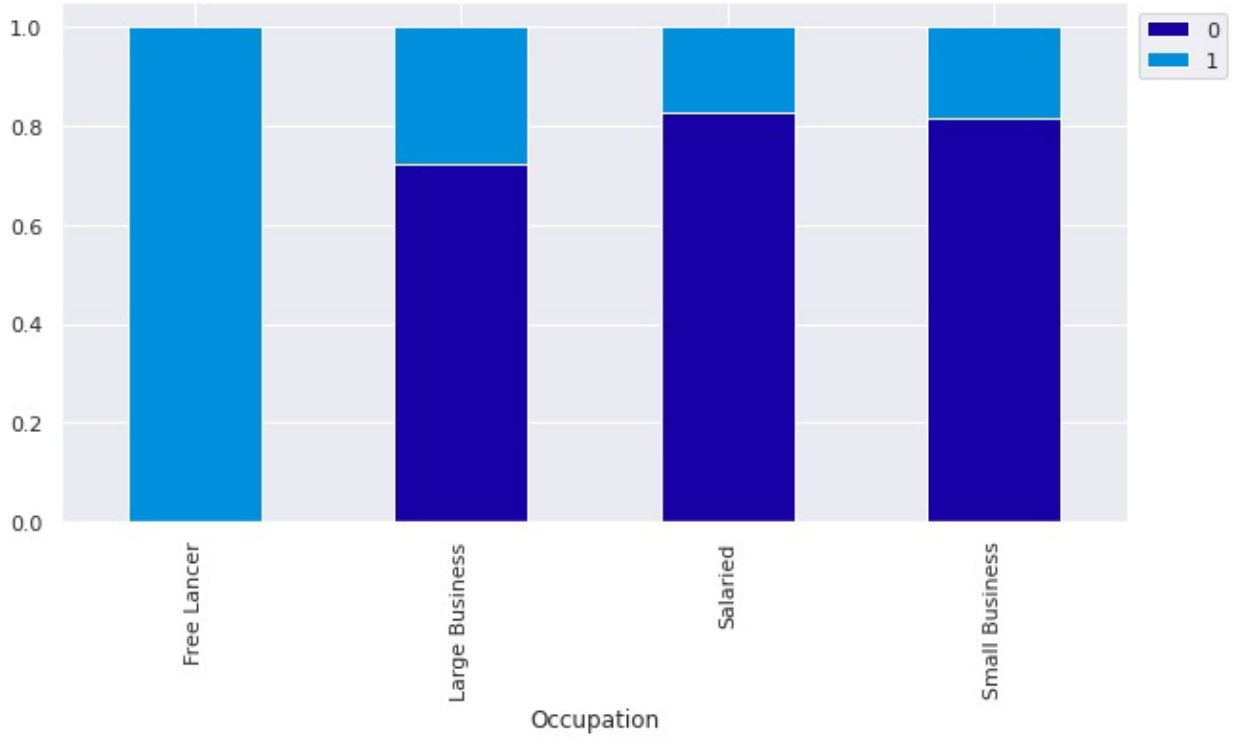




Most are Tier 1. Although more packages were sold in CityTier 1, CityTiers 2 and 3 managed to have a higher percentage in each of their tiers of packages sold.

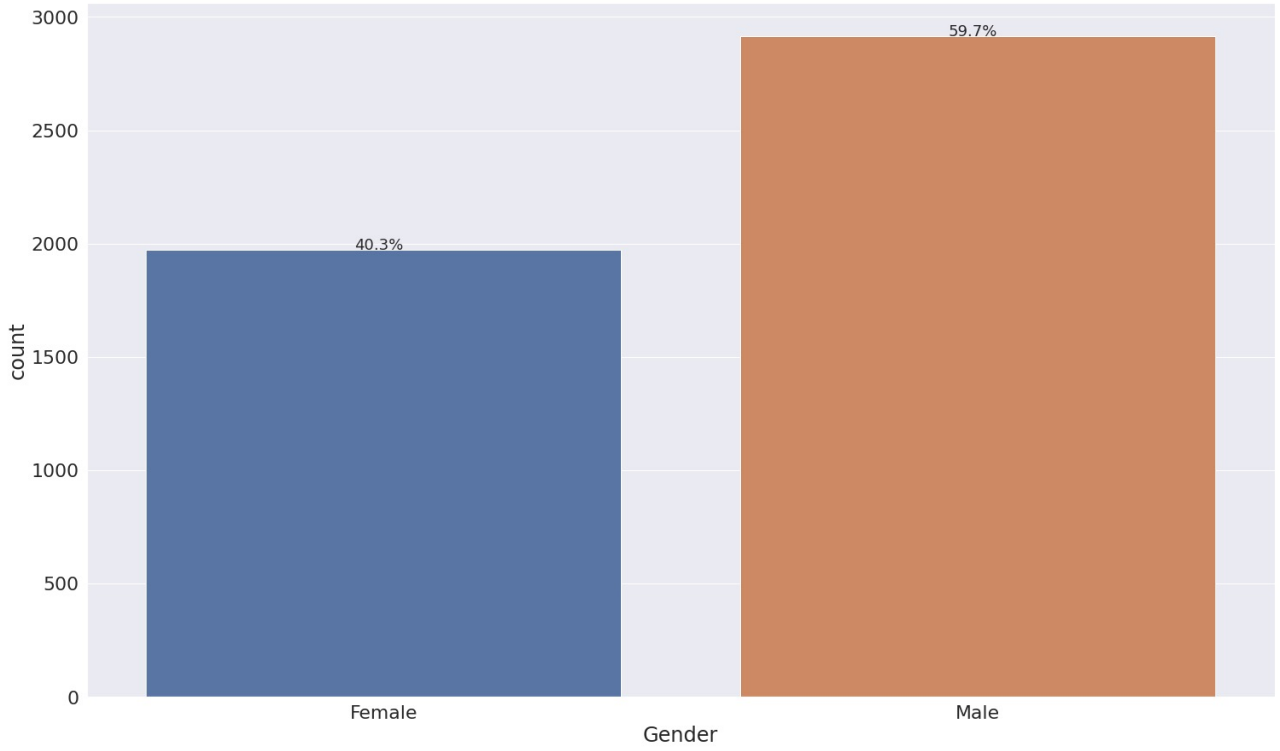
OCCUPATION

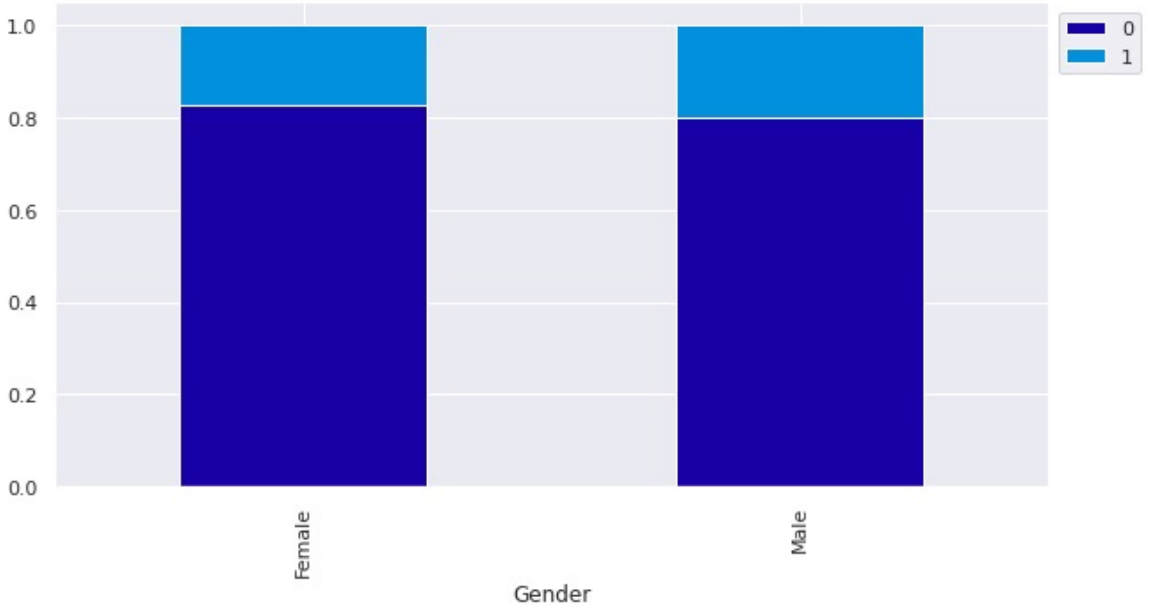




Freelancers make up a very small number of customers but both purchased travel packages. About 28% of Large Business owners purchased packages.

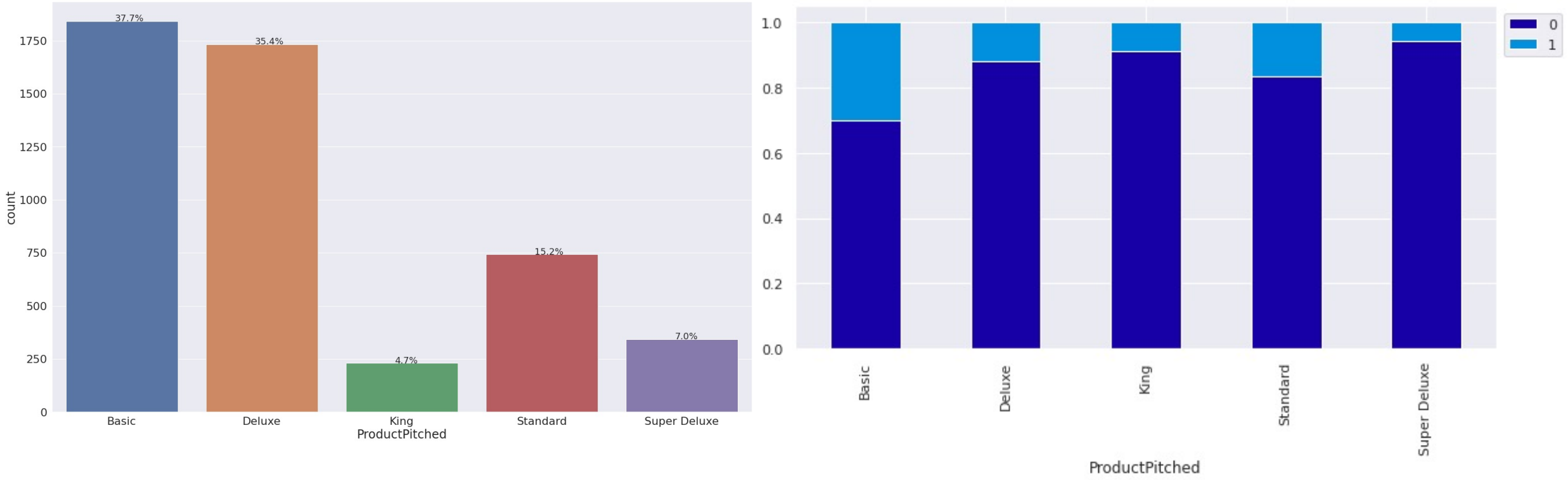
GENDER





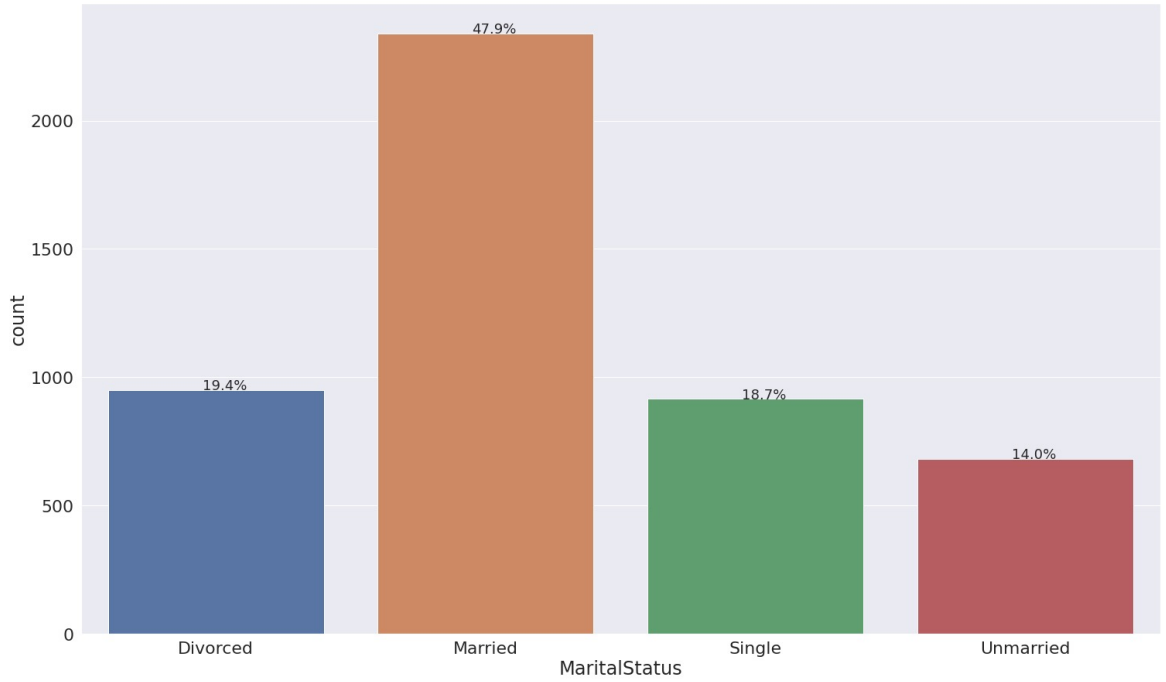
Most customers are male. By a very slight margin, the percentage of males who purchase a package is greater than females.

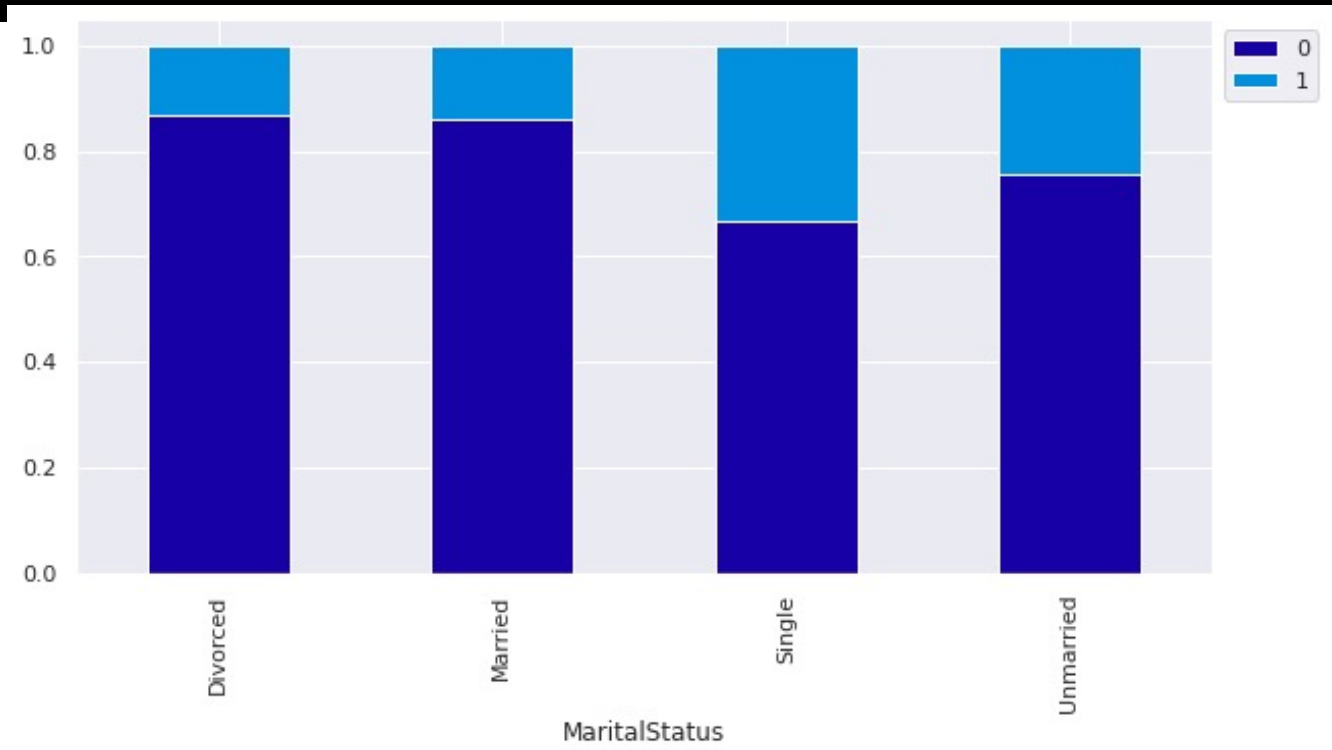
PRODUCT PITCHED



Most salespeople pitched the Basic package. A higher percentage of Basic packages pitched were purchased, with Standard packages being the next in line.

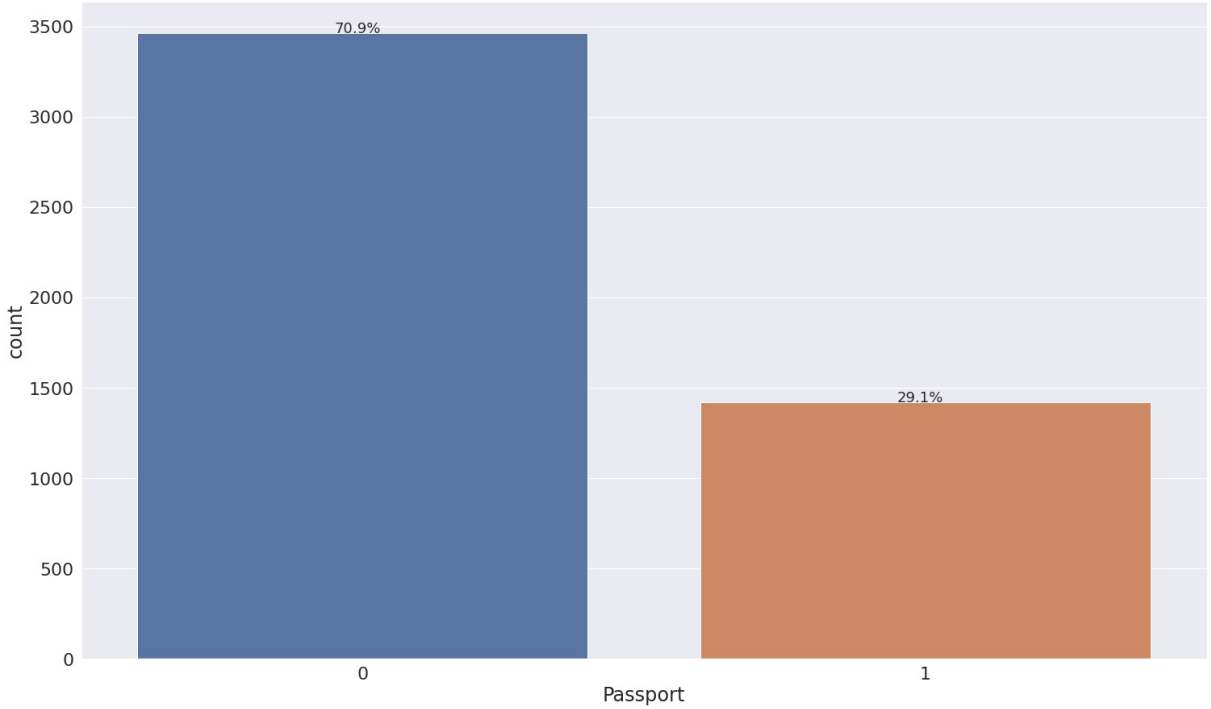
MARITAL STATUS

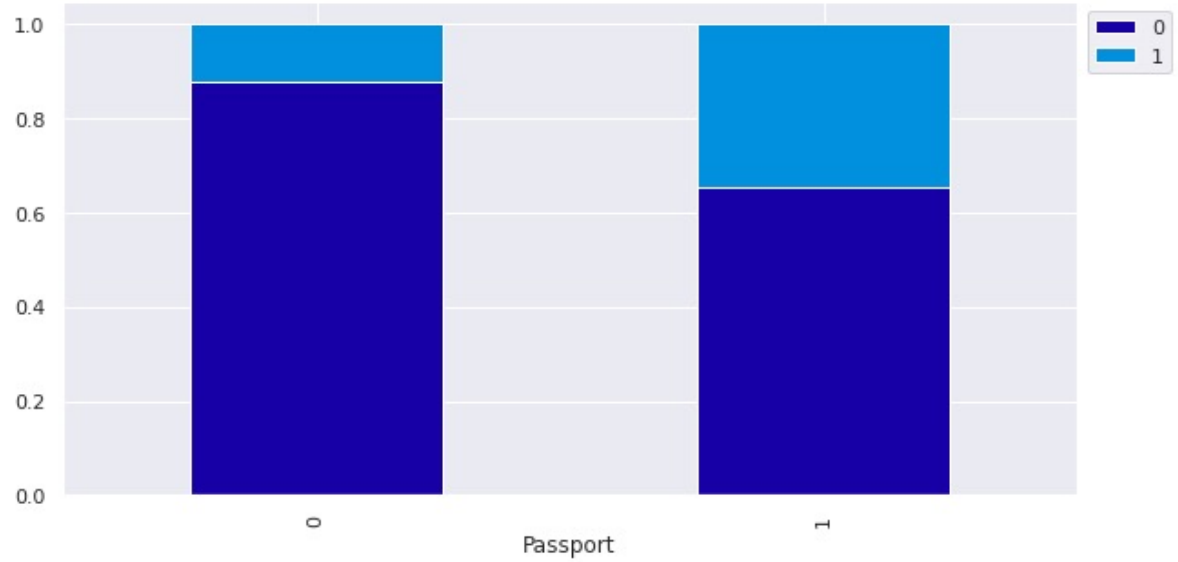




Most customers are married. Although a greater number of packages were sold to Married customers, within the Single category, a larger percentage of those customers decided to purchase a package.

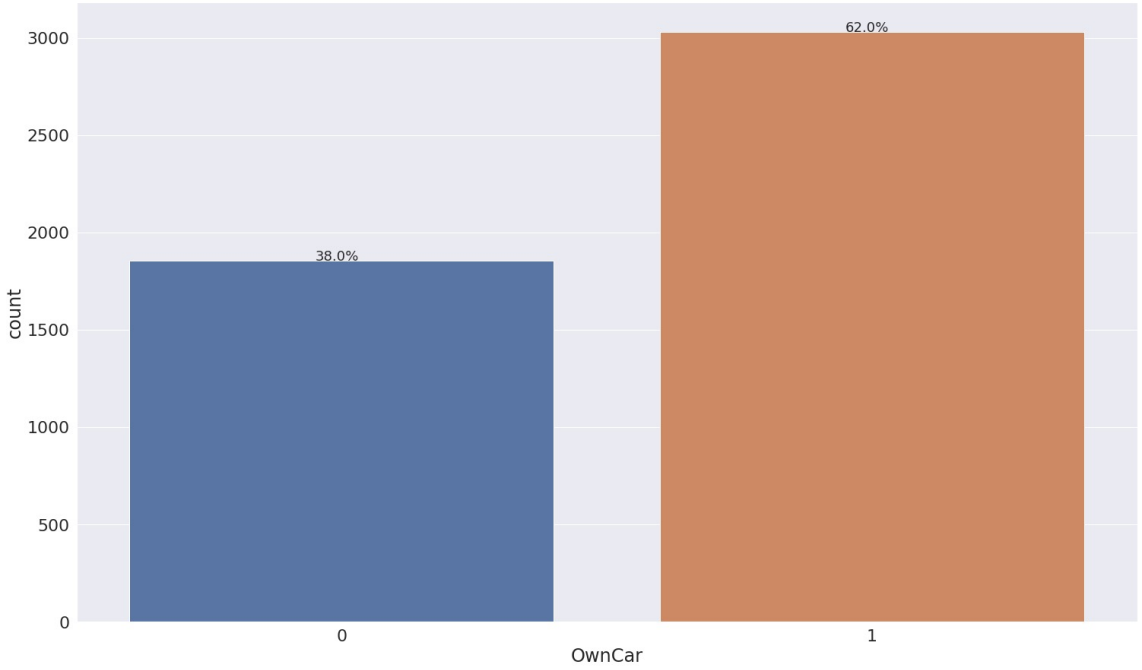
PASSPORT

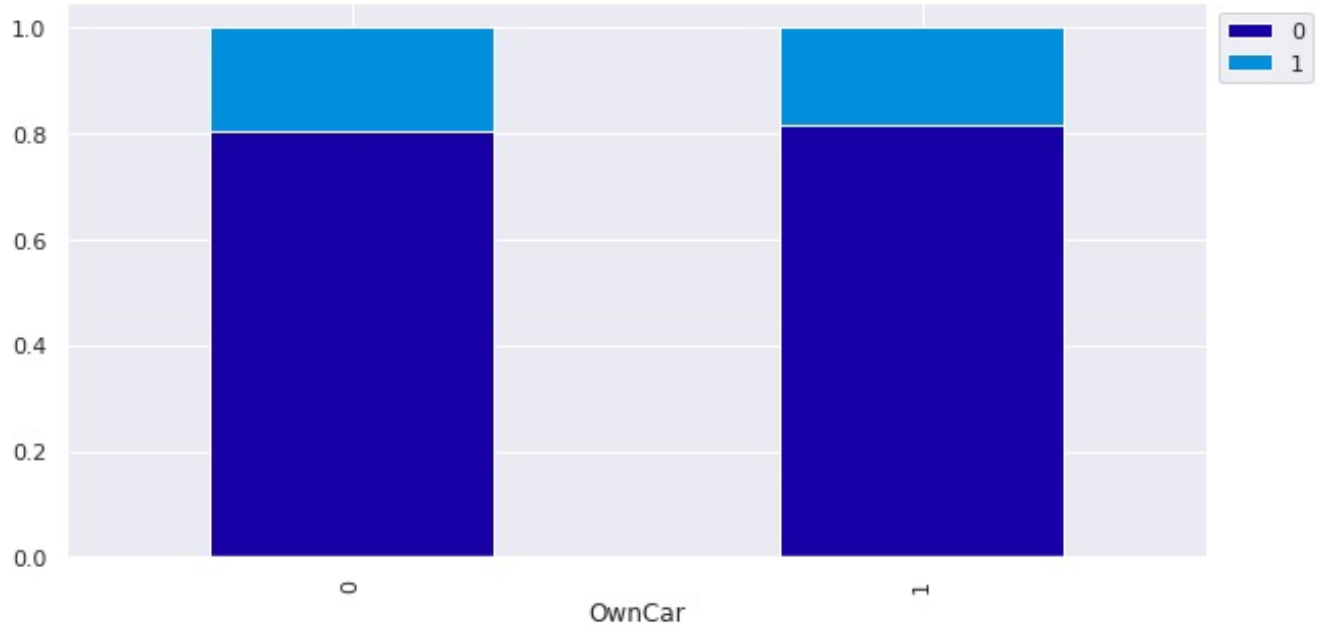




Most customers do not have a passport. A larger percentage of those customers with passports purchased packages.

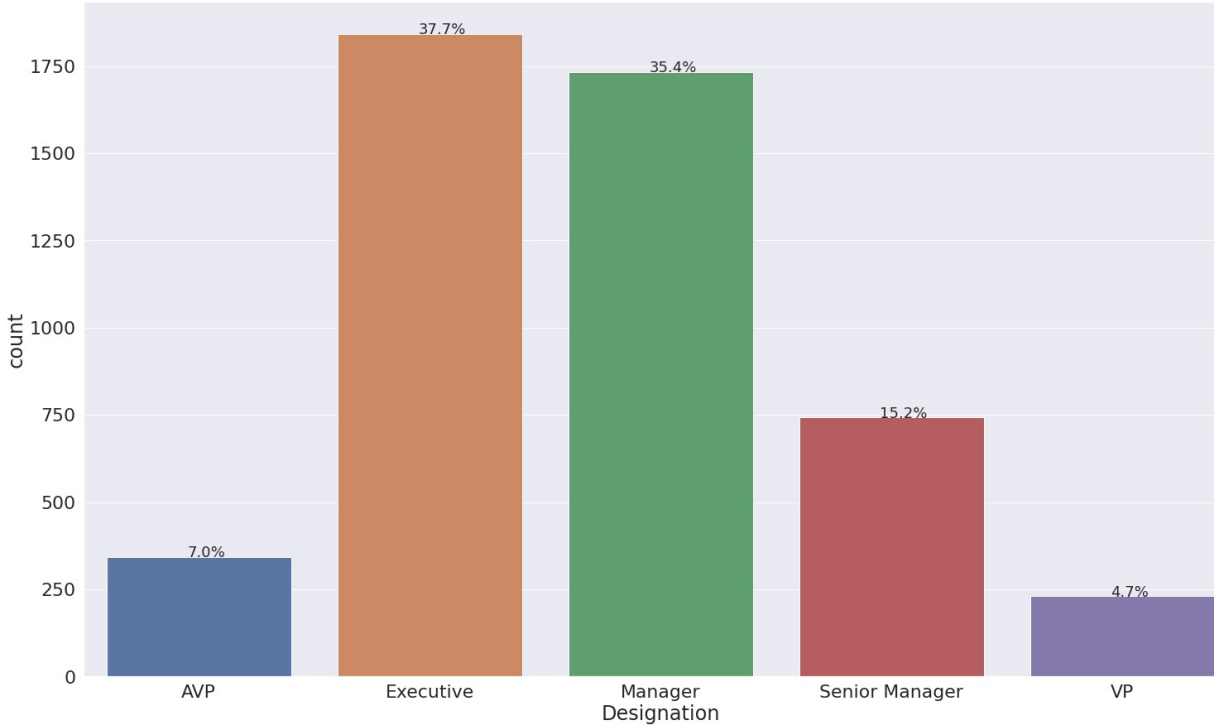
OWN CAR

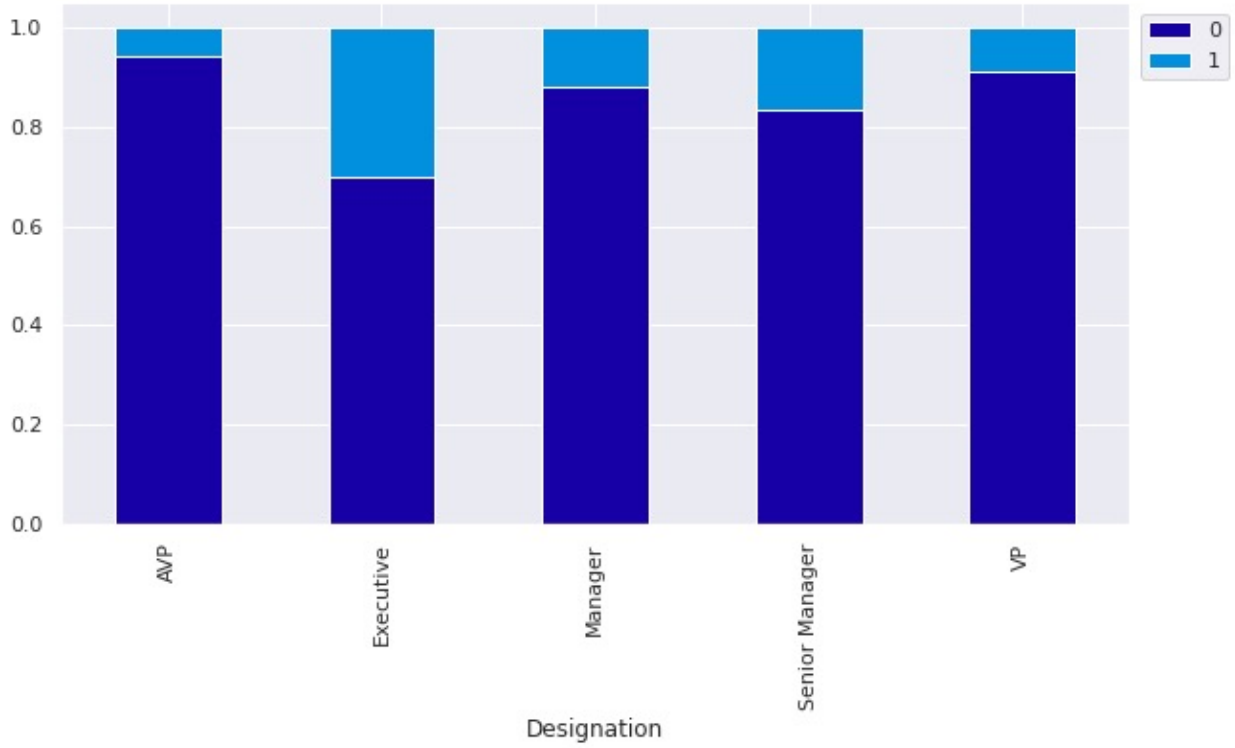




Most customers own their own car. It seems equally likely that a customer will purchase a package whether they own a car or not.

DESIGNATION





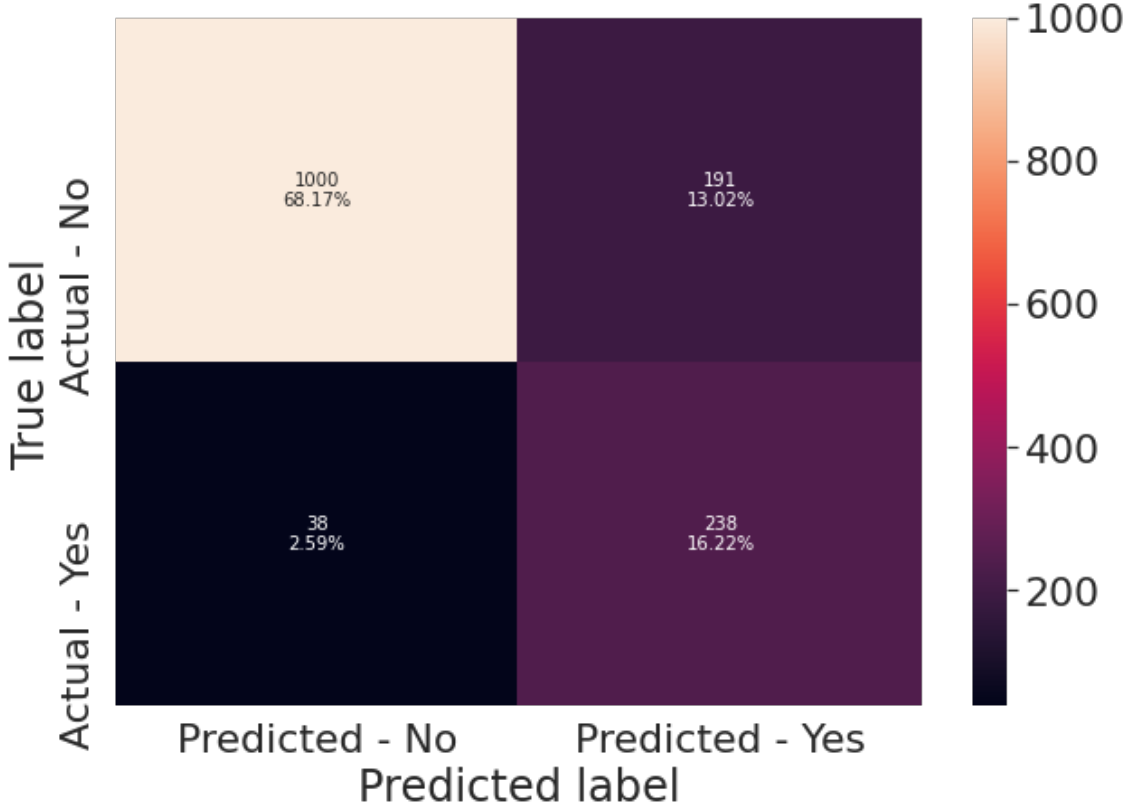
Most are designated as Executive. A larger percentage of the Executive designation purchased a package.

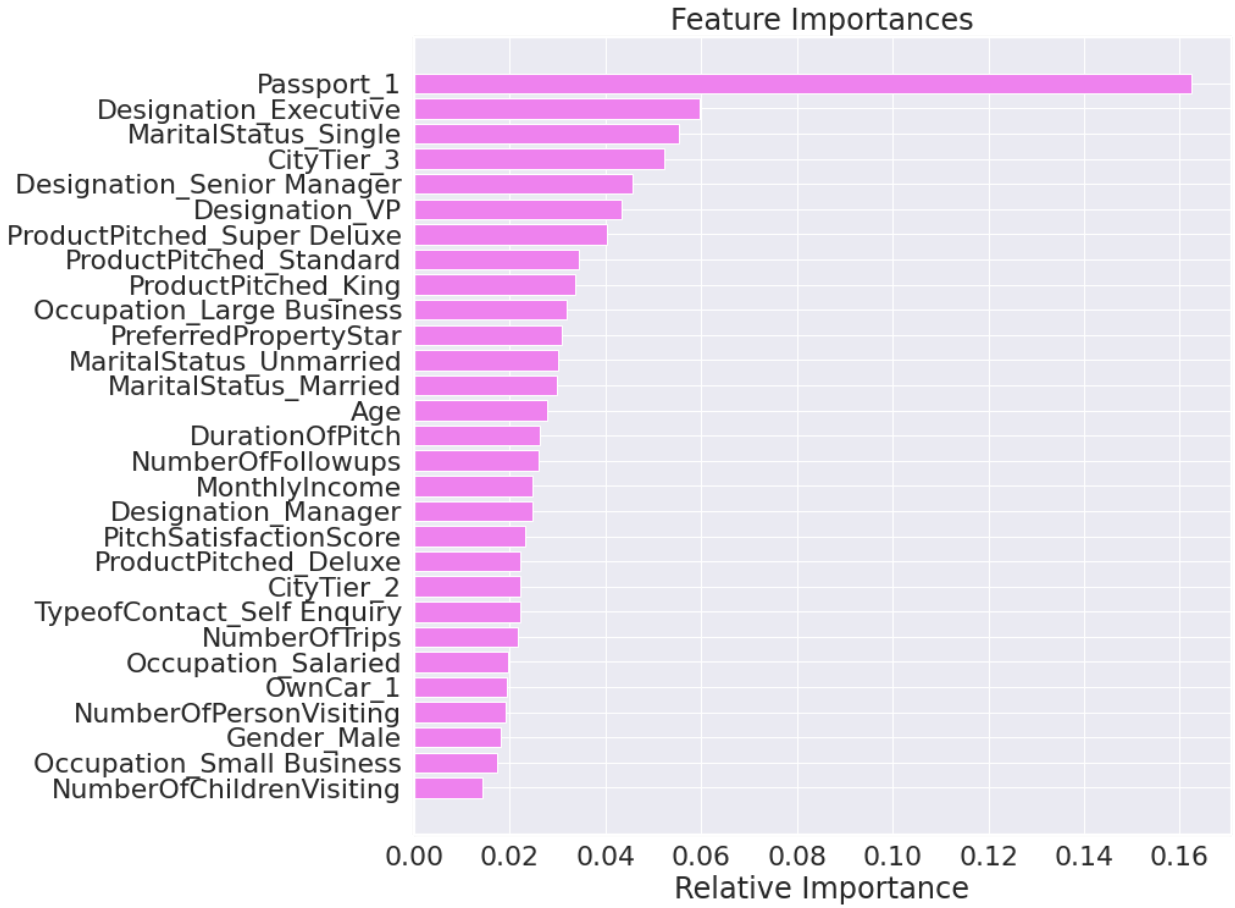
MODEL SUMMARY

	Model	Train_Accuracy	Test_Accuracy	Train_Recall	Test_Recall	Train_Precision	Test_Precision
0	Decision Tree	1.000000	0.890252	1.000000	0.717391	1.000000	0.704626
1	Bagging Classifier	0.995031	0.912065	0.975155	0.630435	0.998410	0.865672
2	Weighted Bagging Classifier	0.995031	0.903885	0.976708	0.554348	0.996830	0.894737
3	Random Forest	1.000000	0.917519	1.000000	0.594203	1.000000	0.947977
4	Weighted Random Forest	1.000000	0.911384	1.000000	0.554348	1.000000	0.956250
5	Tuned Decision Tree	0.750073	0.751875	0.681677	0.688406	0.403122	0.405983
6	Tuned Bagging Classifier	0.679626	0.687798	0.655280	0.695652	0.325617	0.339223
7	Tuned Random Forest	1.000000	0.931152	1.000000	0.692029	1.000000	0.922705
8	AdaBoost	0.845659	0.848671	0.315217	0.311594	0.700000	0.728814
9	Gradient Boost	0.887167	0.869121	0.456522	0.409420	0.890909	0.795775
10	XGBoost	0.999708	0.927062	0.998447	0.692029	1.000000	0.896714
11	Tuned AdaBoost	1.000000	0.884799	1.000000	0.605072	1.000000	0.735683
12	Tuned Gradient Boost	0.946215	0.895024	0.740683	0.557971	0.965587	0.827957
13	Tuned XGBoost	0.892137	0.843899	0.976708	0.862319	0.639878	0.554779
14	Stacking Classifier	0.989477	0.882754	1.000000	0.891304	0.947059	0.634021

The tuned XGBoost model performed the best on the Test data without overfitting.

TUNED XGBOOST MODEL





CONCLUSION

- Passport ranks as by far the most influential predictor. Customers with passports are more likely to purchase a package. Marketing could explore setting up programs to guide potential customers in getting their passports since the vast majority of potential customers do not have passports.
- Designation is an important predictor of whether or not a customer will purchase a package. A higher percentage of Executive members purchase packages. Marketing could explore designing campaigns or packages targeting the Executive class.
- MaritalStatus is an important predictor in determining if a customer will purchase a package. Most packages are sold to married couples, but the single customers had the highest percentage of packages purchased within their segment.
- CityTier is an important predictor in determining whether a customer will purchase a package. The marketing team might further investigate if there are travel trends that dominate certain city tiers (e.g., if CityTier 1 is urban do those customer prefer packages that allow them to travel to more pastoral settings).
- ProductPitched - The packages that seem more successfully pitched are the Basic and Standard. The Marketing team needs to further investigate if this is because of cost or some other characteristic of the package itself (destinations, availability, amenities included, etc.)
- Number of children and persons visiting are not important features for prediction. Marketing need not focus on group or perhaps even family-friendly marketing strategies.
- Further research should be done, perhaps in the form of a survey, to determine whether the planned package for Wellness Tourism is a highly desired package for potential customers as current data does not provide adequate insight.