OYO Rooms Goa

Objective: To perform exploratory data analysis on OYOs in Goa, and predict the rating new & unrated OYOs in Goa

Code For Scraping the Data

```
from selenium import webdriver
from bs4 import BeautifulSoup
import pandas as pd
import os
os.remove("hotel3.csv")
driver = webdriver.Chrome(executable_path=r"C:\Users\dayne\chromedriver.exe") #Set the path to chromedriver
headers = {'User-Agent': 'Mozilla/5.0 (X11; Ubuntu; Linux x86_64; rv:77.0) Gecko/20100101 Firefox/77.0'}
driver.get("https://www.oyorooms.com/111193/")
driver.find_element_by_css_selector('.c-1vevh8c').click()
a=[]
content = driver.page_source
soup = BeautifulSoup(content)
name1=soup.find('h1', attrs={'itemprop':'name'})
a.append(name1.text.strip())
name2=soup.find('div', attrs={'class':'c-1qcdse5'})
a.append(name2.text.strip())
name3=soup.find('div', attrs={'class':'c-v9oteh'})
a.append(name3.text.strip())
name4=soup.find('span', attrs={'itemprop':'streetAddress'})
a.append(name4.text.strip())
name5=soup.find('span', attrs={'class':'listingPrice__finalPrice listingPrice__finalPrice--black'})
a.append(name5.text.strip())
xyz = soup.findAll('div',href=False, attrs={'itemprop':'amenityFeature'})
xyz
for lap in xyz:
    name=lap.find('div', attrs={'itemprop':'name'})
    a.append(name)
df = pd.DataFrame({'Product Name':a})
df.to_csv('hotel3.csv', index=False, encoding='utf-8')
```

Scraping for hotels was done individually on each hotel website, **to obtain data such as Address, Amenities, Price, Name, Category**

Code For Regression

```
#Importing The Libraries
import pandas as pd
import numpy as np
```

#Importing the datasets

```
df=pd.read_csv("OYO-Training.csv")
df1=pd.read_csv("OYO-Testing.csv")
```

#Splitting the datasets

```
X = df.iloc[:, 1:-1].values
y = df.iloc[:, -1].values
X1=df1.iloc[:,1:-1].values

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0)
```

#Multiple Linear Regression

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
np.set_printoptions(precision=2)
print("Multiple Linear Regression MSE")
mean_squared_error(y_test,y_pred)
```

```
Multiple Linear Regression MSE
Out[820]: 3.303322686579006
```

```
#Support Vector Regression
from sklearn.svm import SVR
regressor = SVR(kernel = 'rbf')
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
print("SVR MSE")
mean_squared_error(y_test,y_pred)
```

```
SVR MSE
Out[821]: 1.2098364477027768
```

```
#Decision Tree
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
print("Decision Tree MSE")
mean_squared_error(y_test,y_pred)
```

```
Decision Tree MSE
Out[822]: 0.6116666666666666
```

```
#Random Forest
from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators = 10, random_state = 0)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
```

```
np.set_printoptions(precision=2)
print("Random Forest MSE")
mean_squared_error(y_test,y_pred)
```

```
Random Forest MSE
Out[823]: 0.8451749999999999
```

#Best output obtained for Decision Tree Hence we will use it for predicting the ratings of new OYO hotels

```
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(X_train, y_train)
y_pred2=regressor.predict(X1)
a=pd.DataFrame(y_pred2)
a
a.to_csv("oyo-output1.csv")
```

```
Out[831]:
     0
0  4.2
1  4.3
2  4.3
3  4.2
4  4.2
5  4.2
6  3.5
7  4.5
8  2.9
9  4.5
```

Exploratory Data Analysis After Combining The Results

#Importing The Libraries

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

#Adding the combined dataset

```
df=pd.read_csv("OYO-EDA.csv")
```

#Exploratory Data Analysis

```
df['Price'].describe()
df['Rating'].describe()
```

```
In [751]: df['Rating'].describe()
Out[751]:
count    69.000000
mean      4.156087
std       0.316233
min       3.500000
25%       3.920000
50%       4.200000
75%       4.400000
max       4.800000
Name: Rating, dtype: float64
```

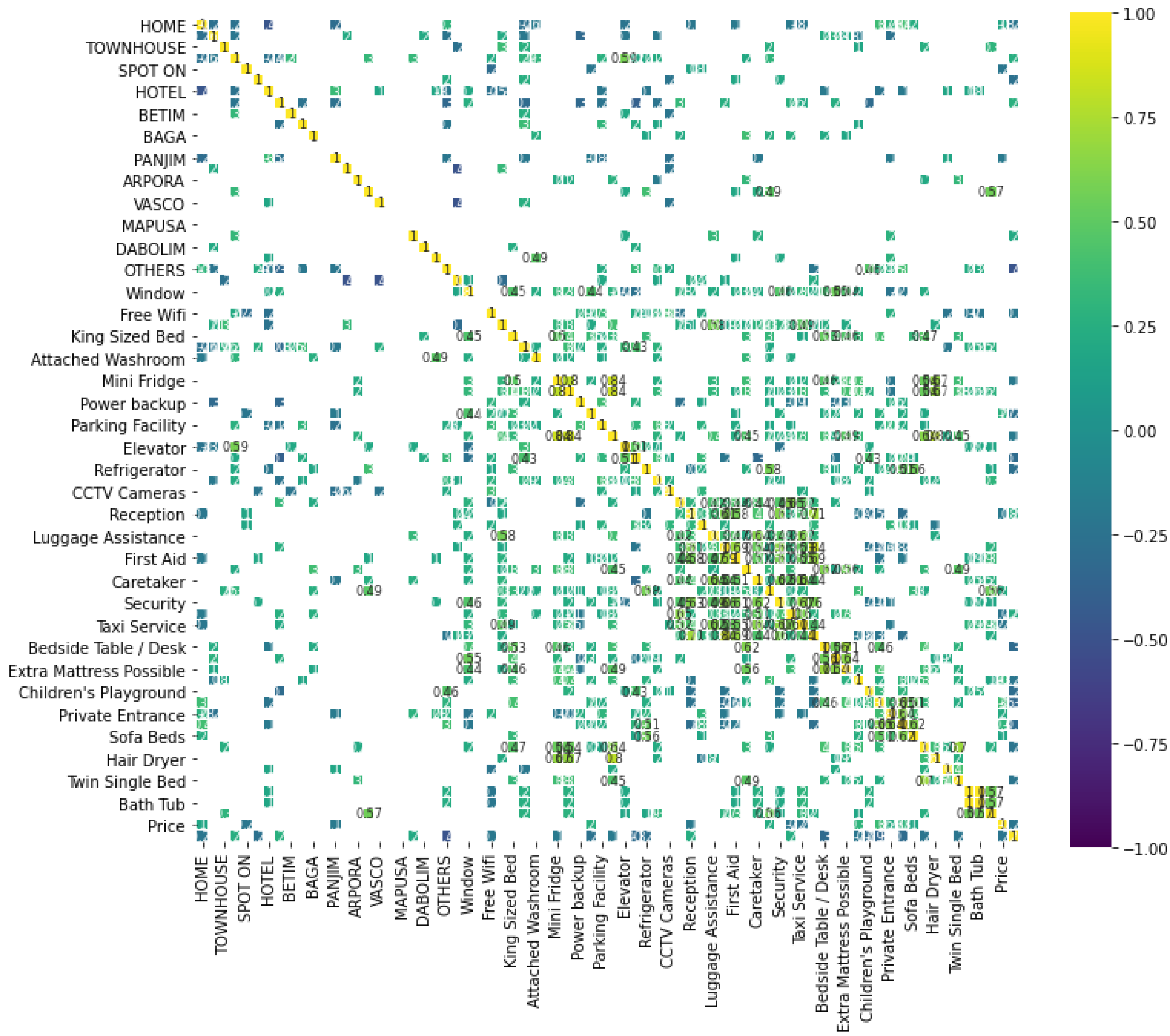
```
In [750]: df['Price'].describe()
Out[750]:
count       69.000000
mean      5537.043478
std       2654.742712
min        827.000000
25%       3825.000000
50%       5110.000000
75%       6827.000000
max      13541.000000
Name: Price, dtype: float64
```

#Generating the correlations
df_num_corr = df.corr()['Price'][1:]
golden_features_list = df_num_corr[abs(df_num_corr) > 0.2].sort_values(ascending=False)
print("There is {} strongly correlated values with SalePrice:\n{}".format(len(golden_features_list), golden_features_list))

```
There is 11 strongly correlated values with SalePrice:
Price                1.000000
Modern Wardrobe      0.362651
Balcony              0.303867
Shower Head          0.300436
Geyser               0.266937
Garden/Backyard      0.231740
Private Entrance     0.225538
SPOT ON             -0.215098
Taxi Service        -0.220191
Car Rentals         -0.222147
Rating              -0.223718
Name: Price, dtype: float64
```

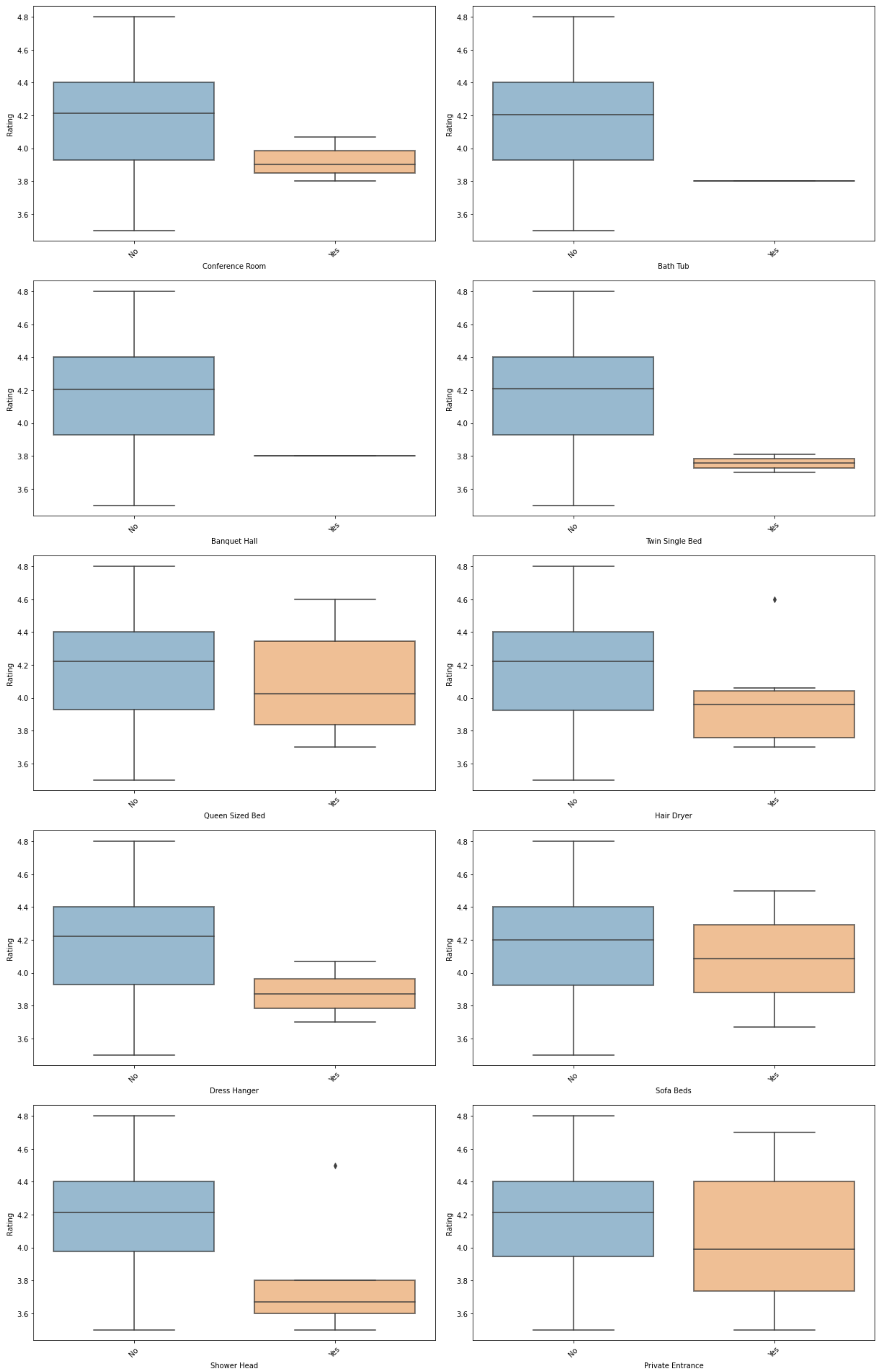
#Heat Map Of Correlations
corr = df.drop(['Name'], axis=1).corr()
plt.figure(figsize=(12, 10))

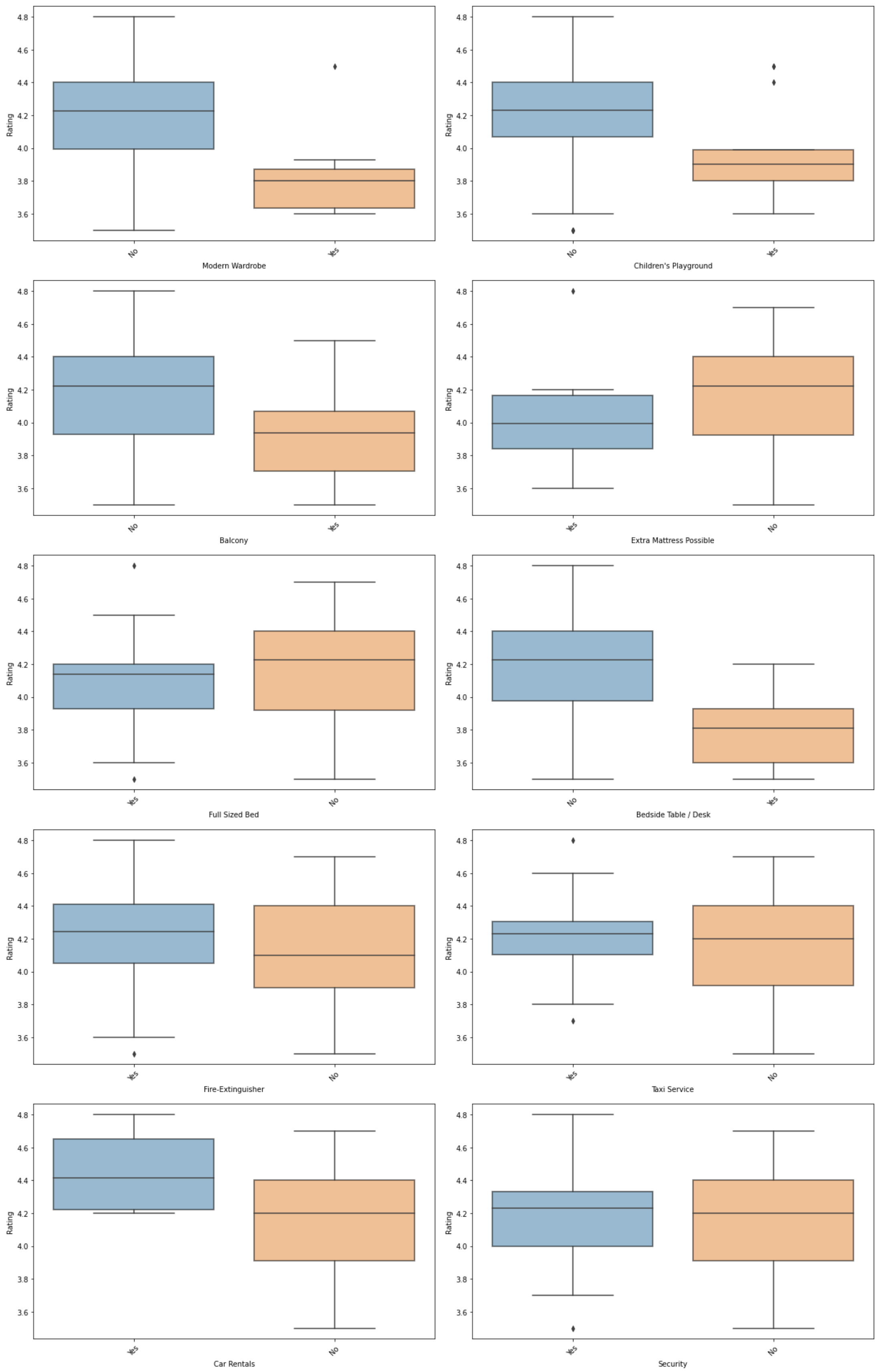
sns.heatmap(corr[(corr >= 0.175) | (corr <= -0.175)],
    cmap='viridis', vmax=1.0, vmin=-1.0, linewidths=0.1,
    annot=True, annot_kws={"size": 8}, square=True);



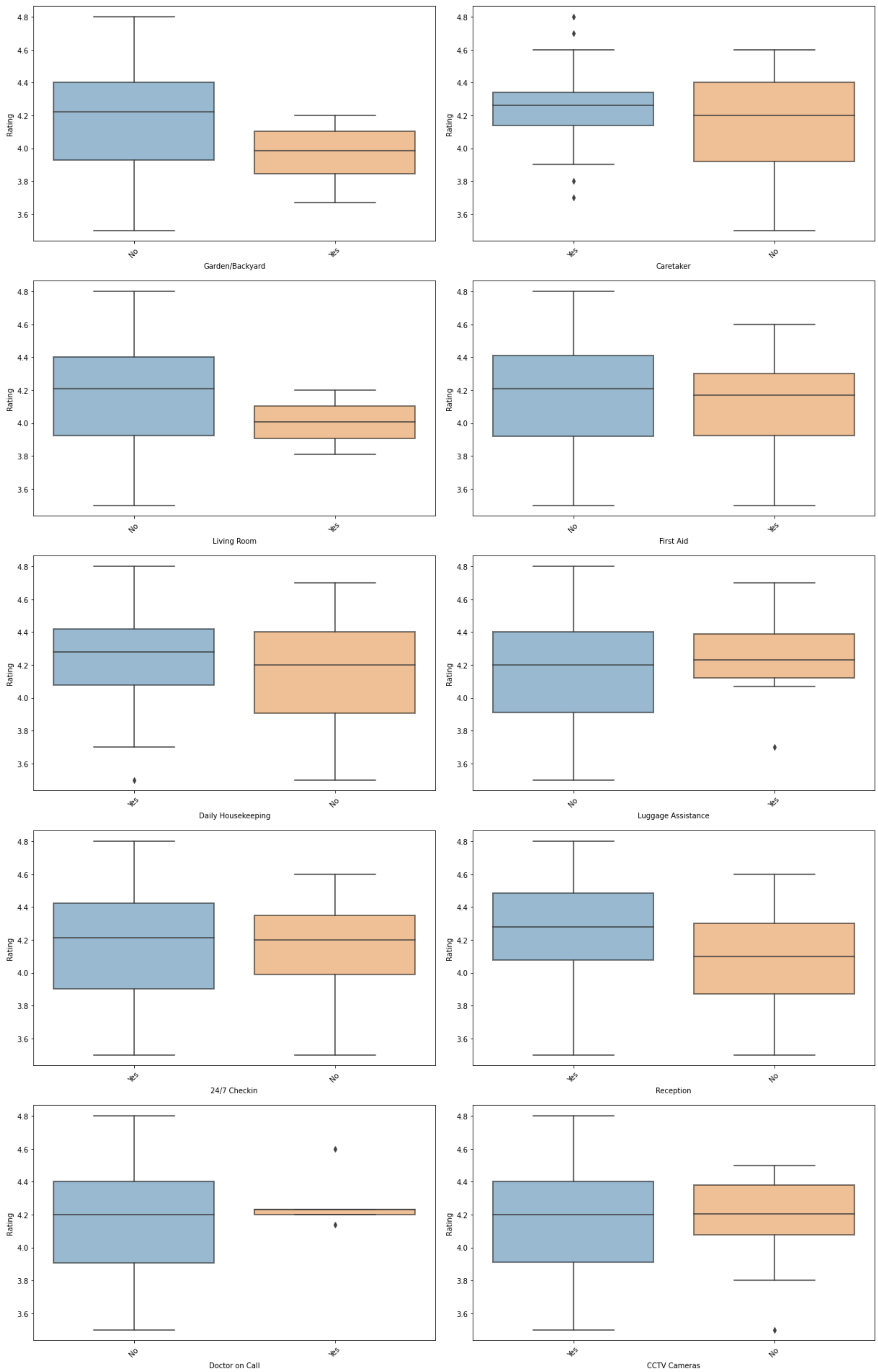
#Creating Box-Plots with Y axis as Rating

```
df_categ2=df.drop(['Name'],axis=1)
df_categ2=df_categ.replace(0,"No")
df_categ2=df_categ.replace(1,"Yes")
for i in range(72):
    plt.figure(figsize = (10, 6))
    ax = sns.boxplot(x=df_categ2.iloc[:,i], y='Rating', data=df_categ2)
    plt.setp(ax.artists, alpha=.5, linewidth=2, edgecolor="k")
    plt.xticks(rotation=45)
```

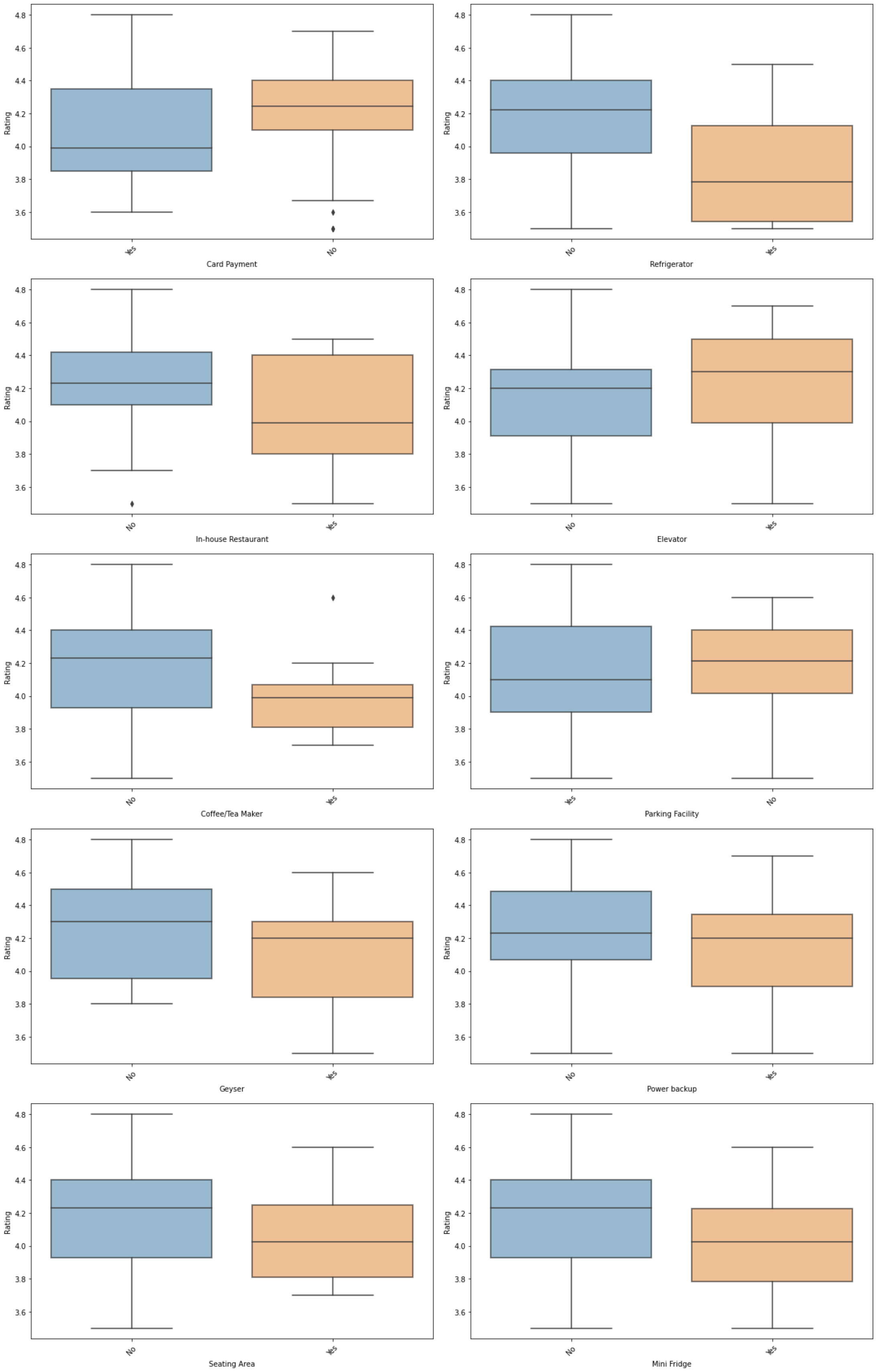

Rating
4.8
4.6
4.4
4.2
4.0
3.8
3.6
No
Yes
Conference Room
Bath Tub
Banquet Hall
Twin Single Bed
Queen Sized Bed
Hair Dryer
Dress Hanger
Sofa Beds
Shower Head
Private Entrance

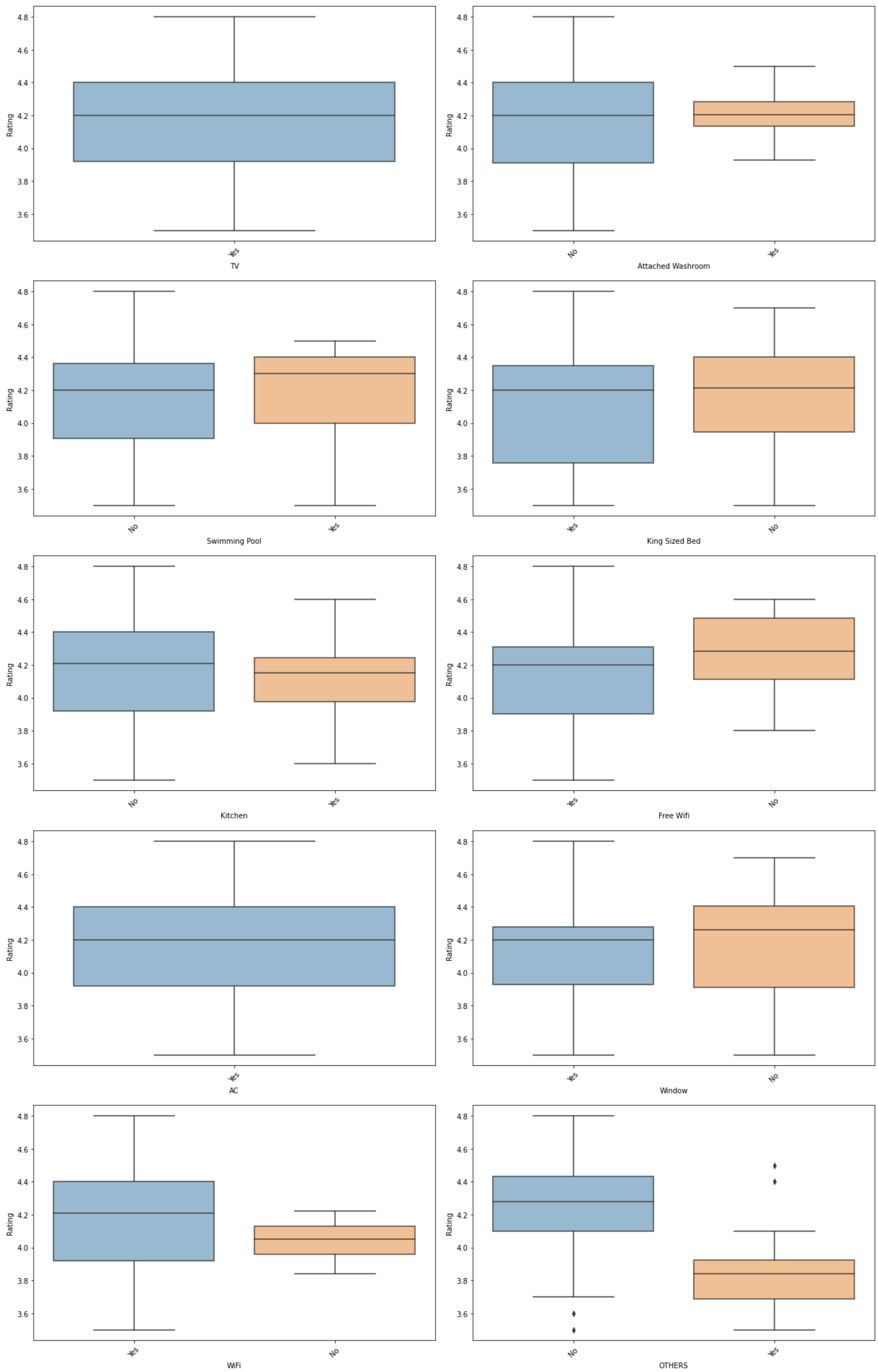


Rating
4.8
4.6
4.4
4.2
4.0
3.8
3.6
No
Yes
Modern Wardrobe
Children's Playground
Balcony
Extra Mattress Possible
Full Sized Bed
Bedside Table / Desk
Fire-Extinguisher
Taxi Service
Car Rentals
Security

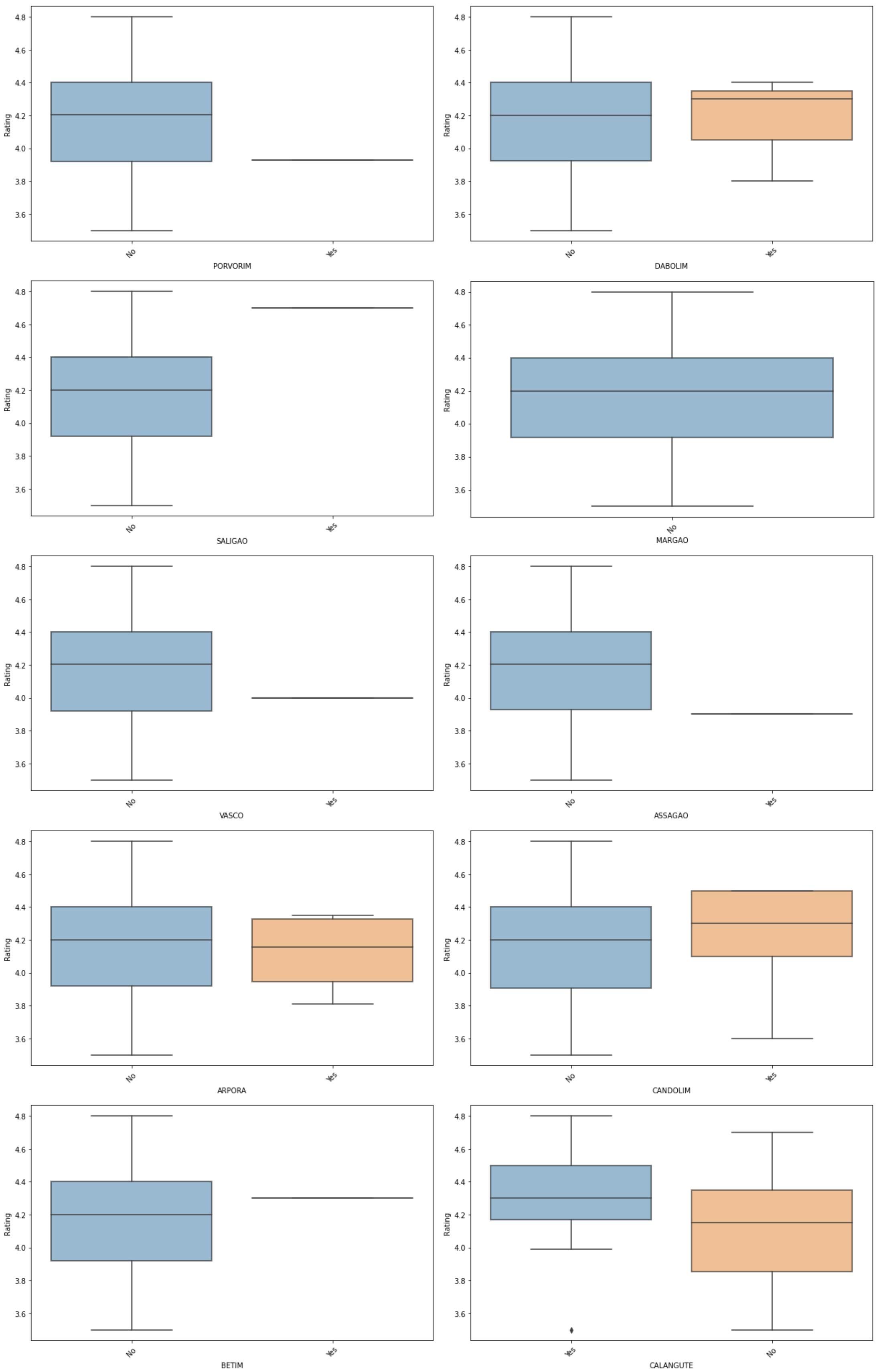


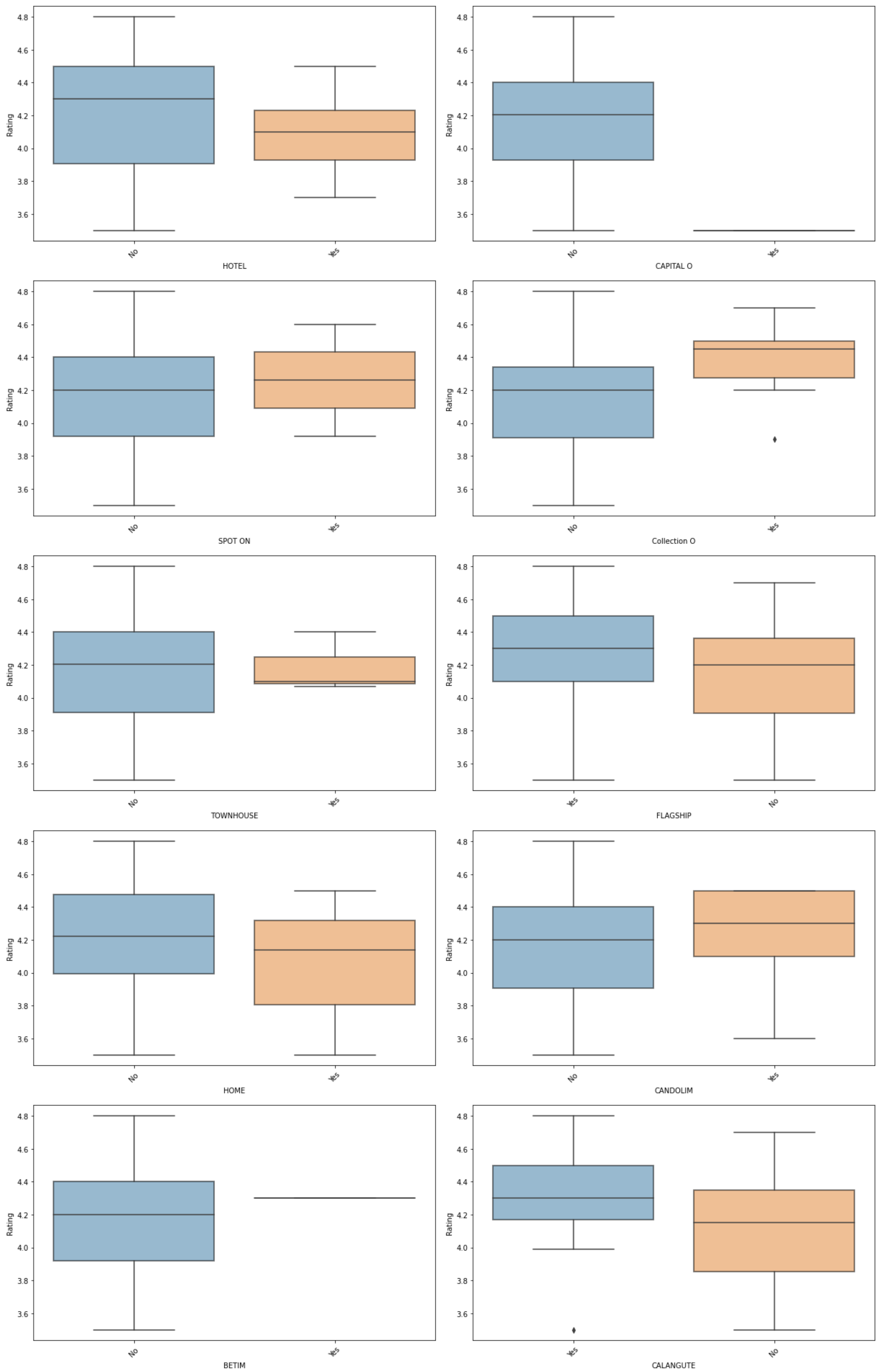
Rating
Garden/Backyard
No
Yes
Caretaker
Yes
No
Living Room
No
Yes
First Aid
No
Yes
Daily Housekeeping
Yes
No
Luggage Assistance
No
Yes
24/7 Checkin
Yes
No
Reception
Yes
No
Doctor on Call
No
Yes
CCTV Cameras
Yes
No
3.6
3.8
4.0
4.2
4.4
4.6
4.8

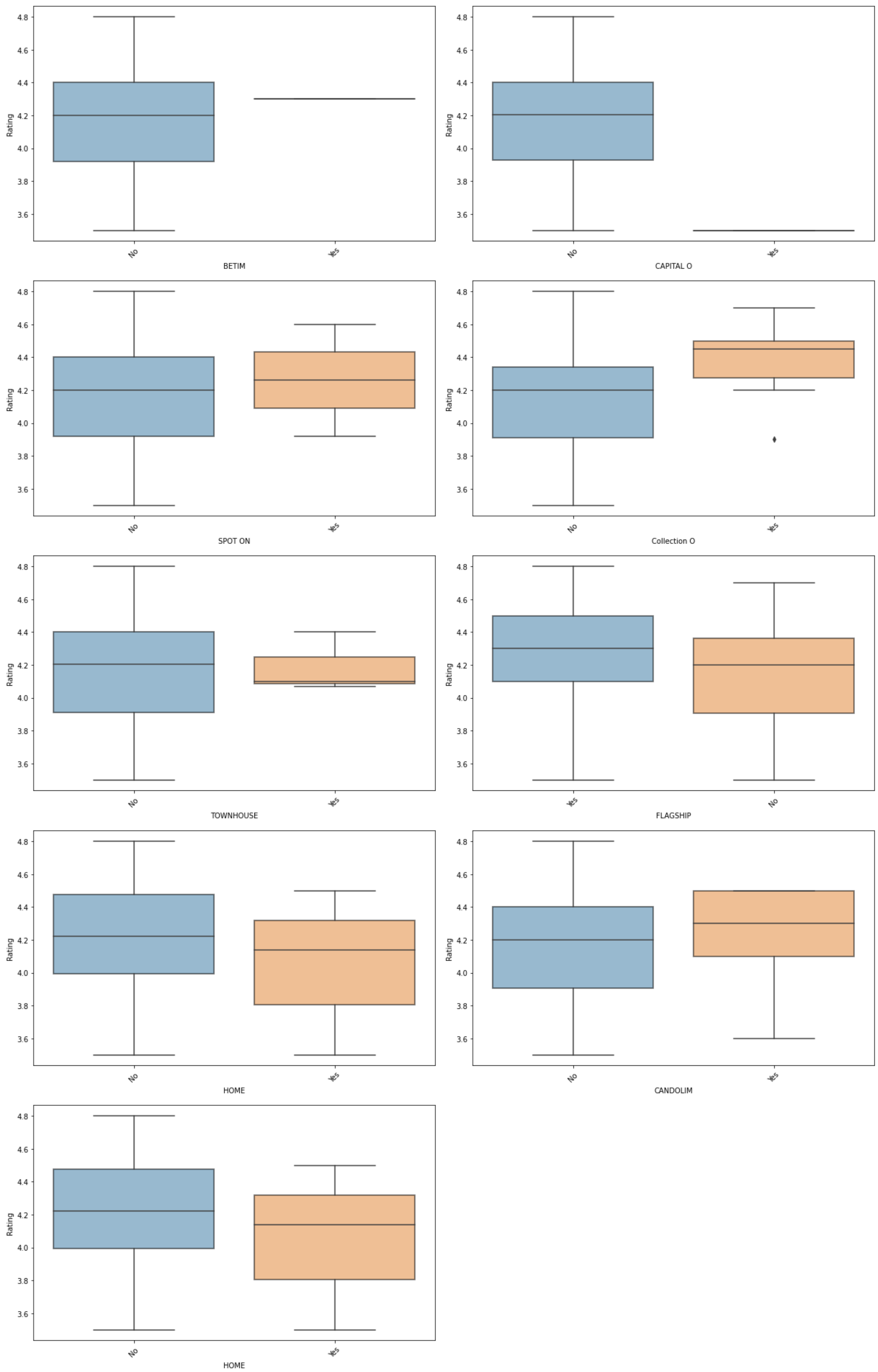

Rating
Yes
No
Card Payment
No
Yes
Refrigerator
No
Yes
In-house Restaurant
No
Yes
Elevator
No
Yes
Coffee/Tea Maker
Yes
No
Parking Facility
No
Yes
Geyser
No
Yes
Power backup
No
Yes
Seating Area
No
Yes
Mini Fridge



Rating
TV
Attached Washroom
Swimming Pool
King Sized Bed
Kitchen
Free Wifi
AC
Window
WiFi
OTHERS
Yes
No

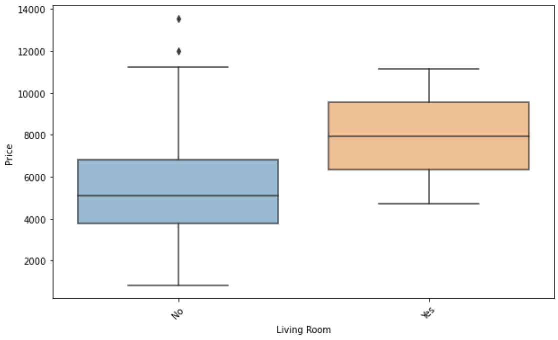

PORVORIM
DABOLIM
SALIGAO
MARGAO
VASCO
ASSAGAO
ARPORA
CANDOLIM
BETIM
CALANGUTE
Rating
No
Yes

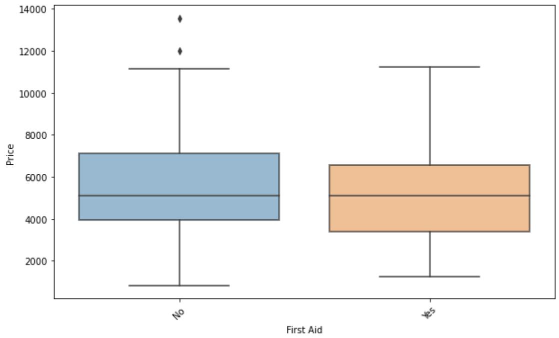

Rating
4.8
4.6
4.4
4.2
4.0
3.8
3.6
No
Yes
HOTEL
CAPITAL O
SPOT ON
Collection O
TOWNHOUSE
FLAGSHIP
HOME
CANDOLIM
BETIM
CALANGUTE

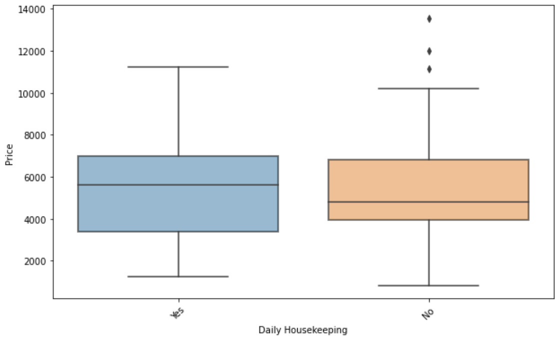


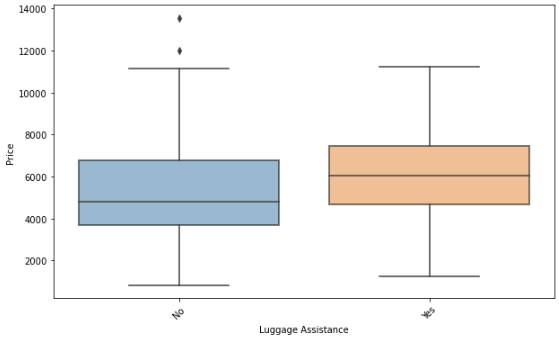
#Boxplots With Y axis as Price

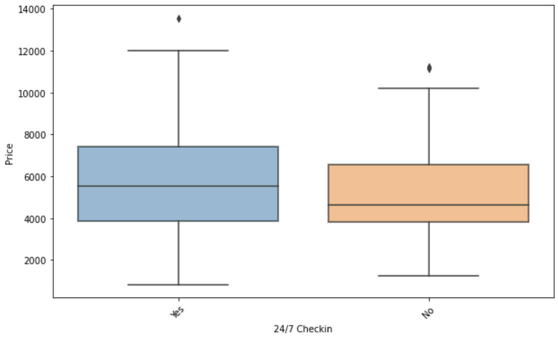
```
for i in range(72):
    plt.figure(figsize = (10, 6))
    ax = sns.boxplot(x=df_categ2.iloc[:,i], y=df['Price'], data=df_categ2)
    plt.setp(ax.artists, alpha=.5, linewidth=2, edgecolor="k")
    plt.xticks(rotation=45)
```

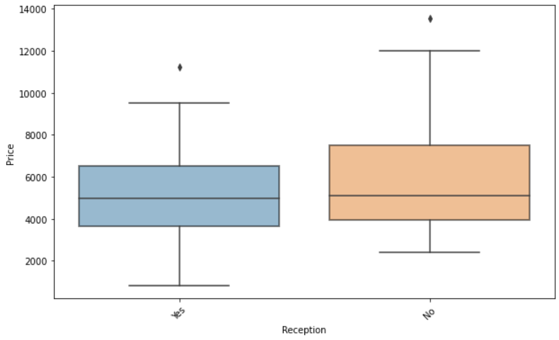


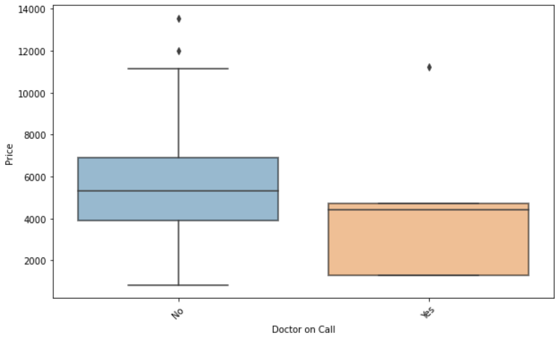


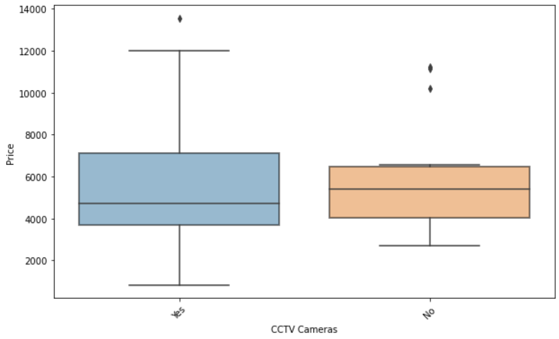


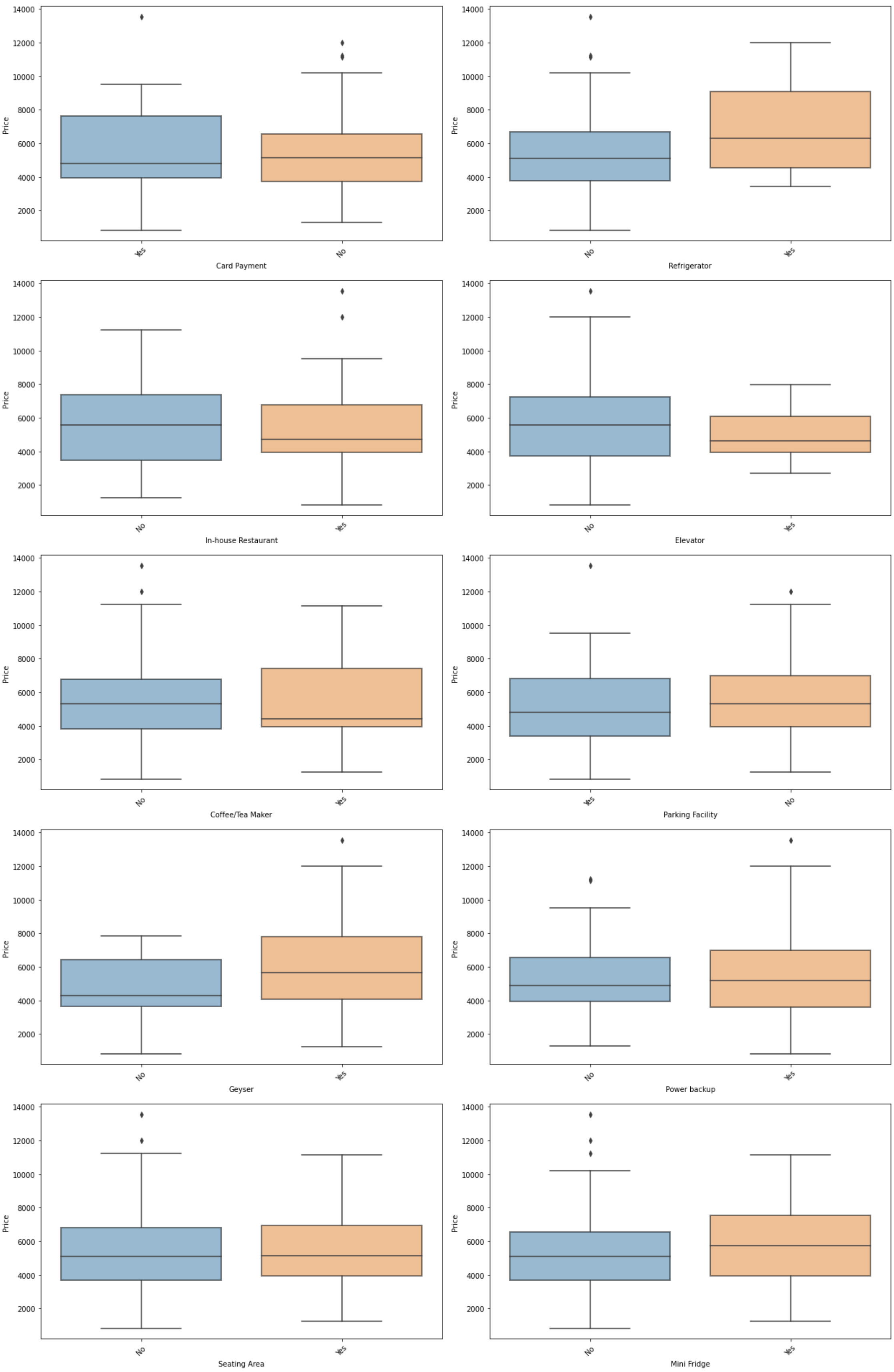


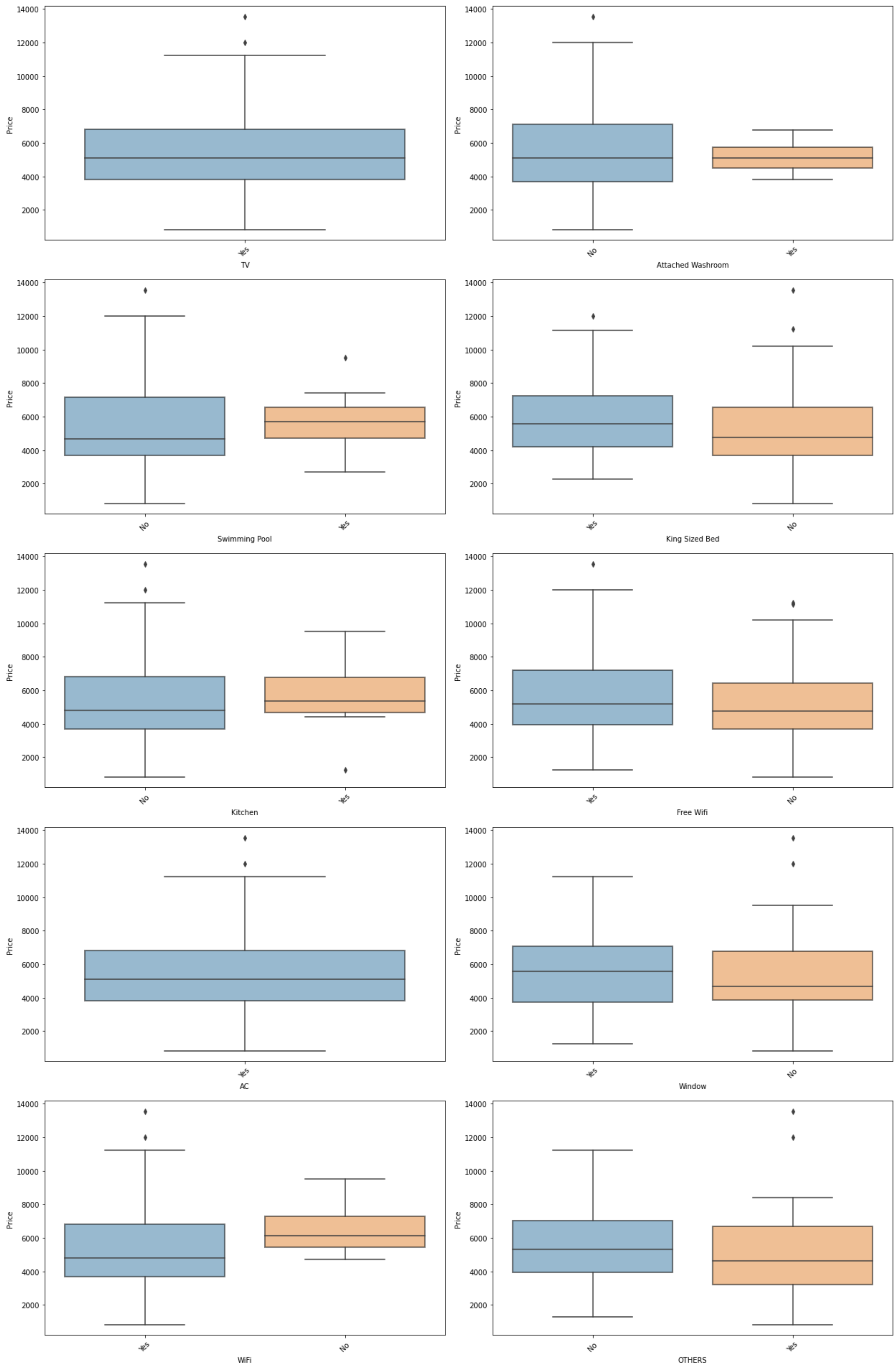




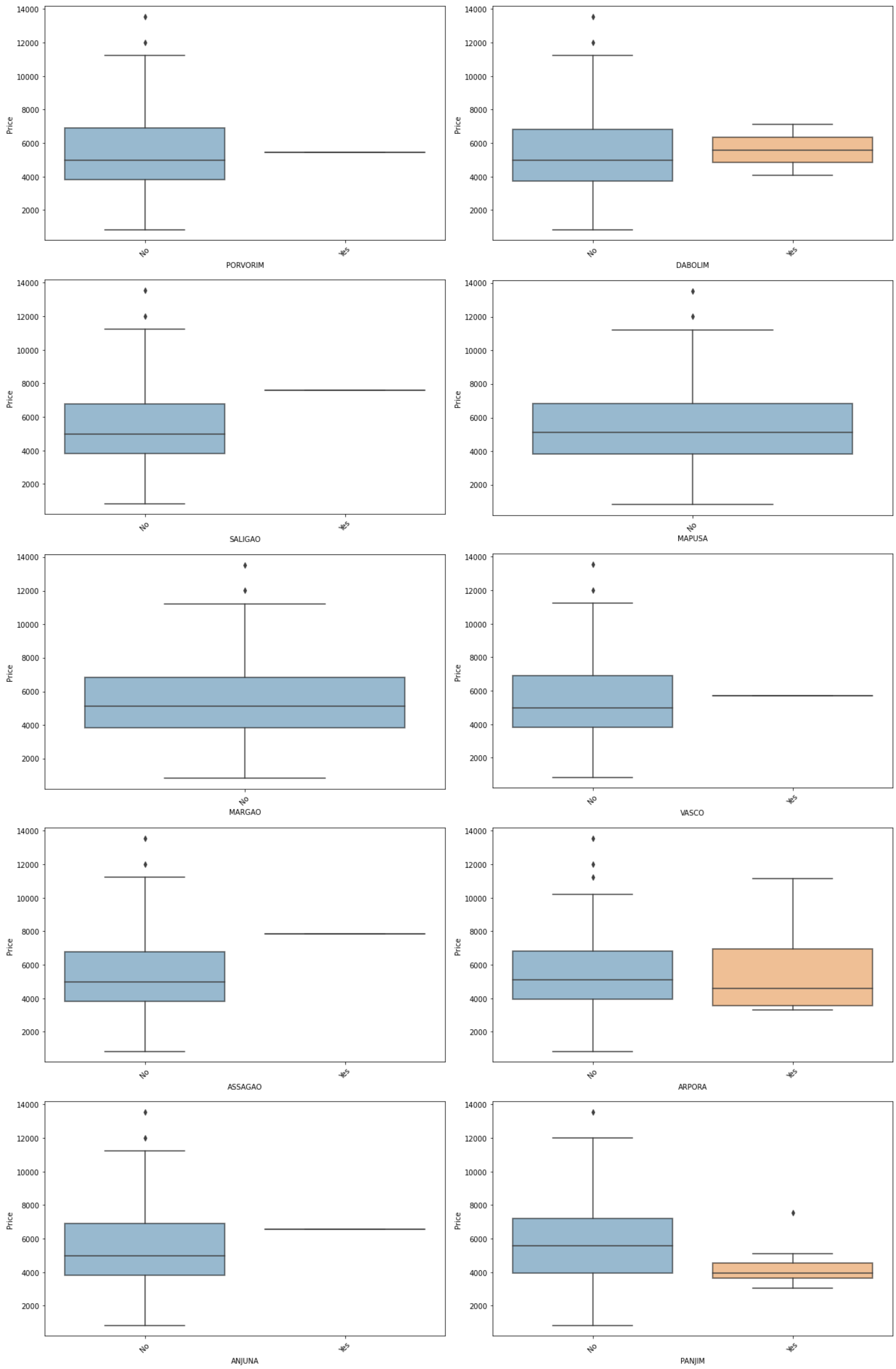




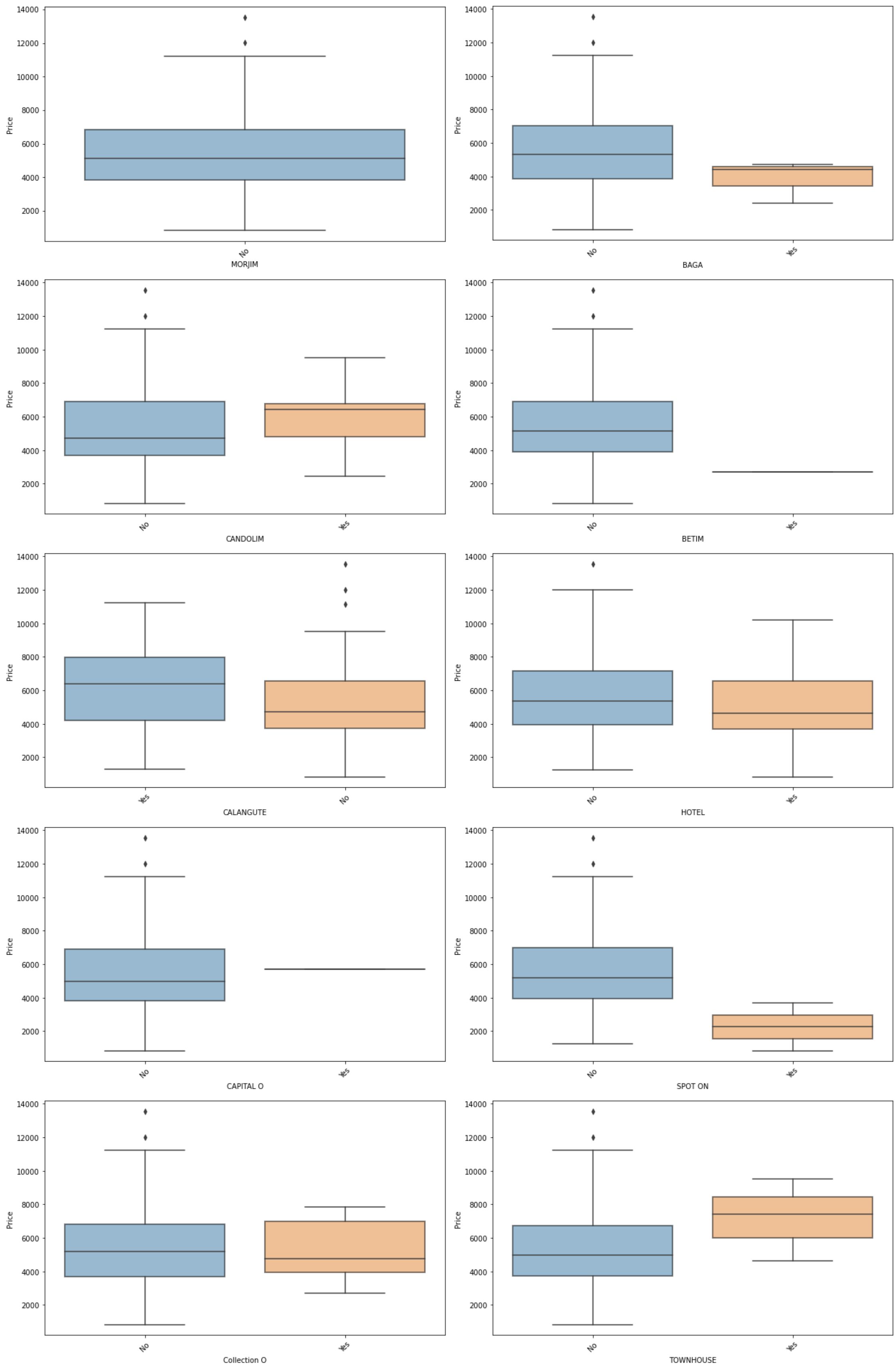

Price
14000
12000
10000
8000
6000
4000
2000
Yes
No
Card Payment
Refrigerator
In-house Restaurant
Elevator
Coffee/Tea Maker
Parking Facility
Geyser
Power backup
Seating Area
Mini Fridge

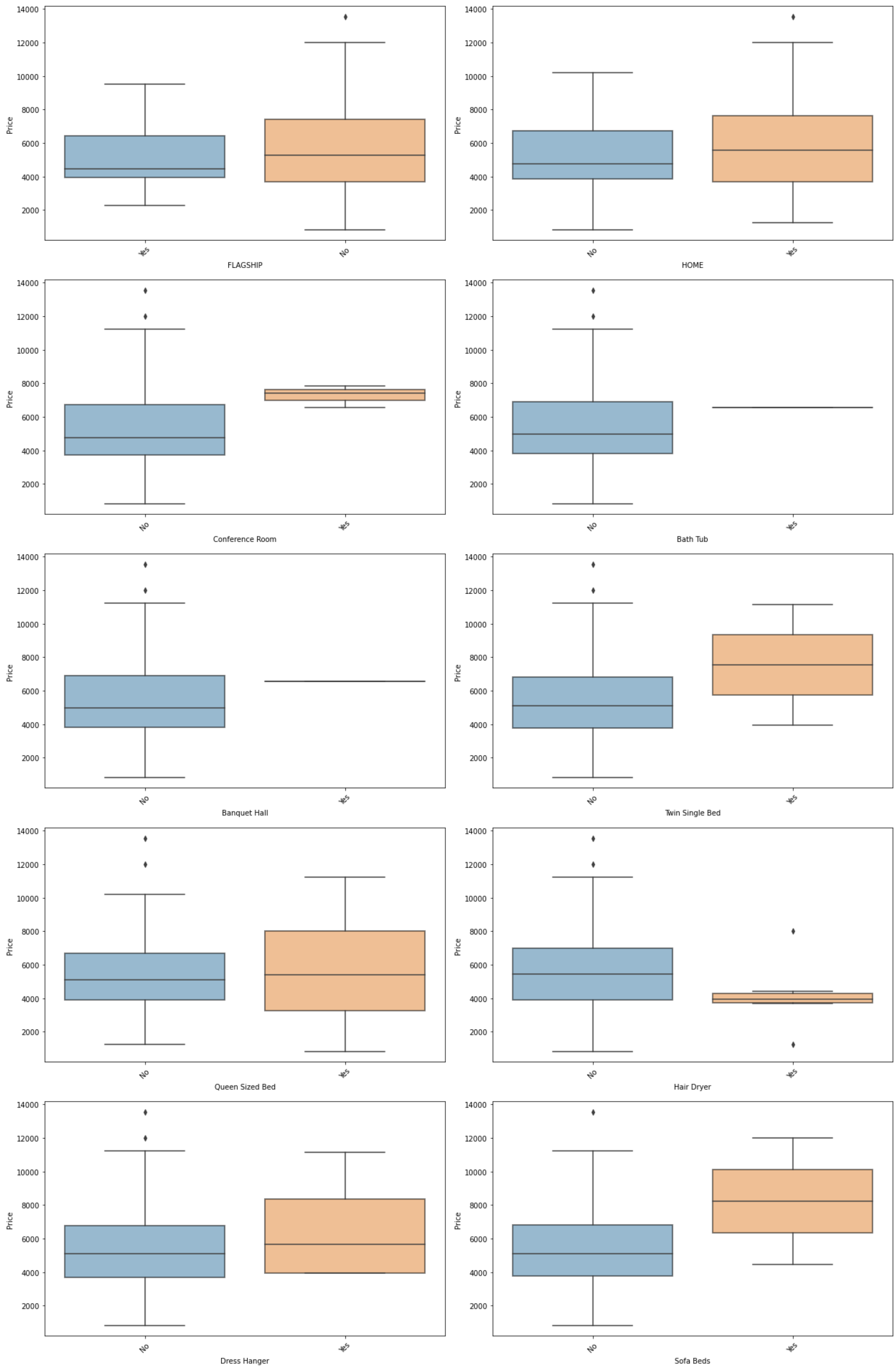


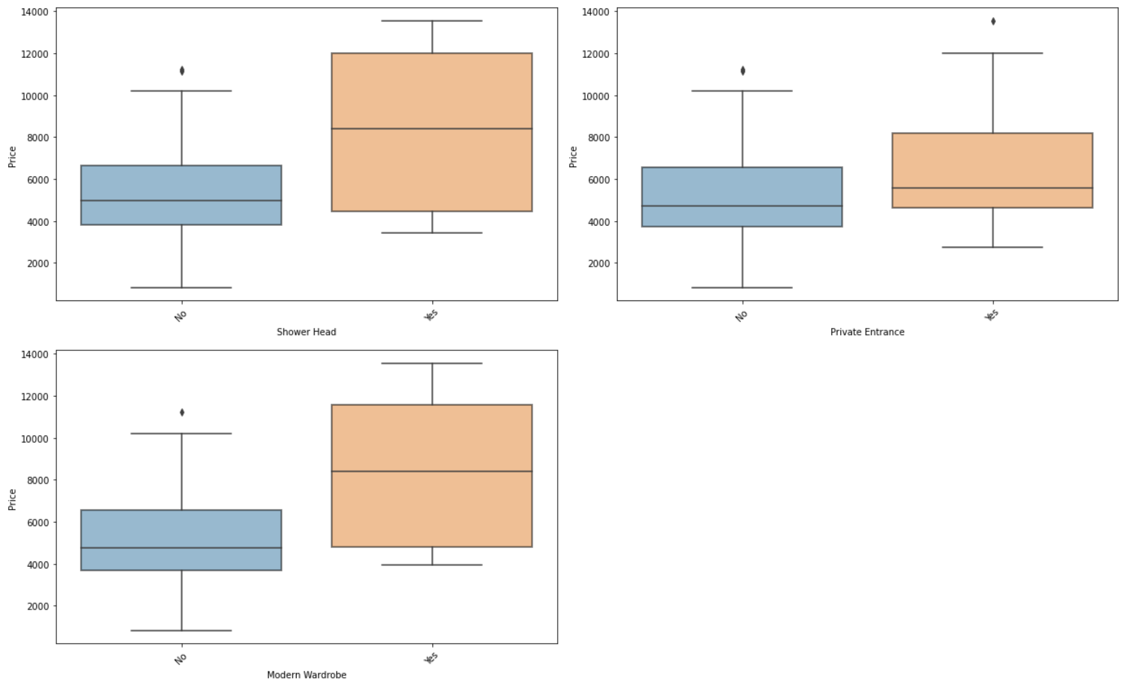
14000
12000
10000
8000
6000
4000
2000
Price
Yes
TV
No
Yes
Attached Washroom
No
Yes
Swimming Pool
Yes
No
King Sized Bed
No
Yes
Kitchen
Yes
No
Free Wifi
Yes
AC
Yes
No
Window
Yes
No
WiFi
No
Yes
OTHERS

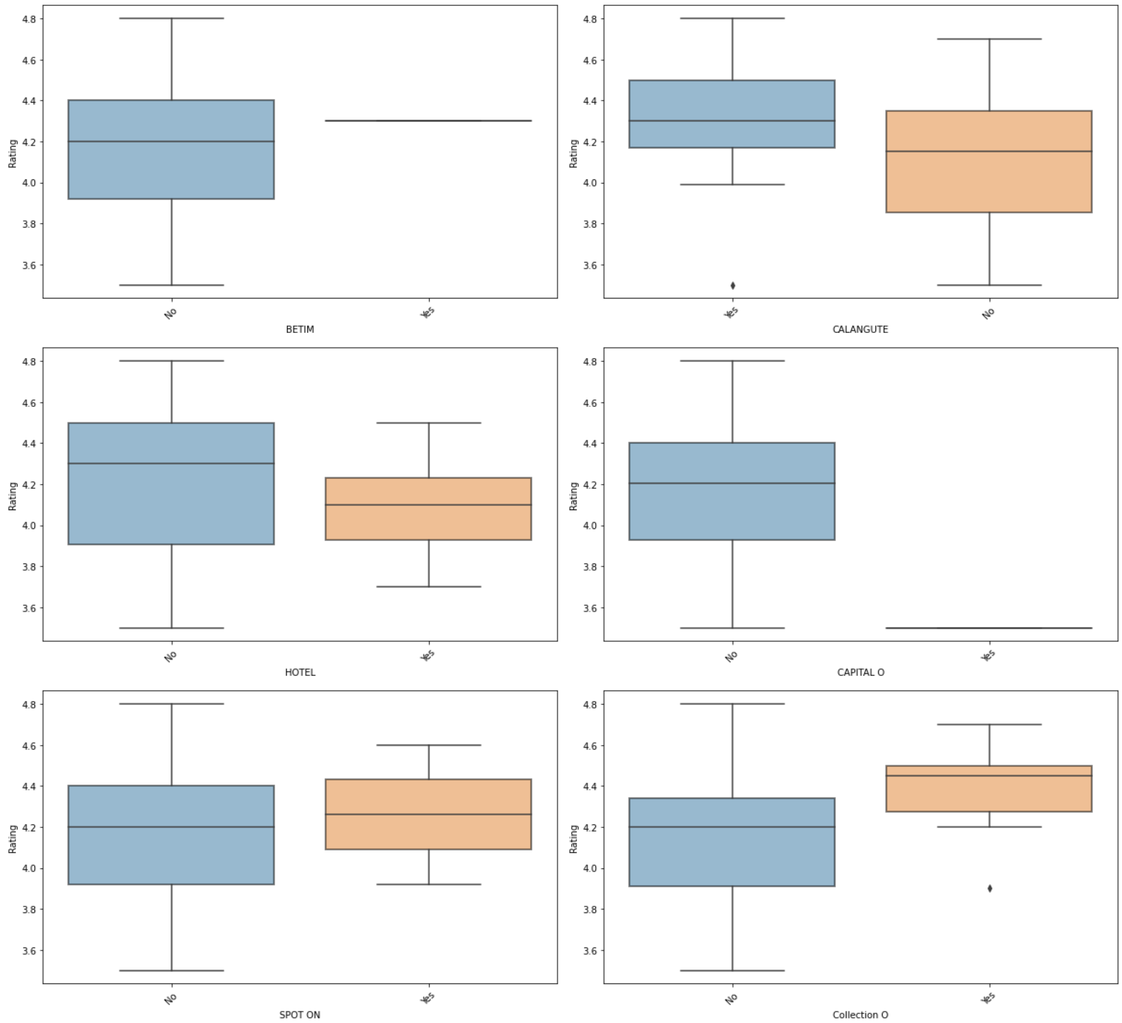


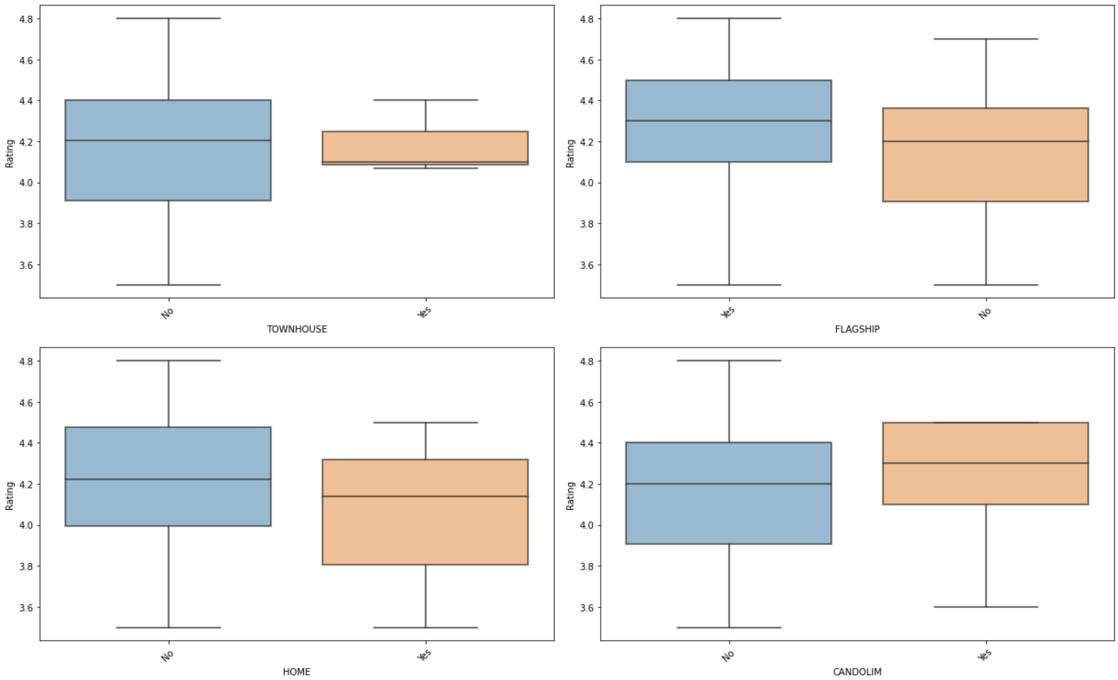
Price
14000
12000
10000
8000
6000
4000
2000
No
Yes
PORVORIM
DABOLIM
SALIGAO
MAPUSA
MARGAO
VASCO
ASSAGAO
ARPORA
ANJUNA
PANJIM


Price
MORJIM
BAGA
CANDOLIM
BETIM
CALANGUTE
HOTEL
CAPITAL O
SPOT ON
Collection O
TOWNHOUSE
No
Yes

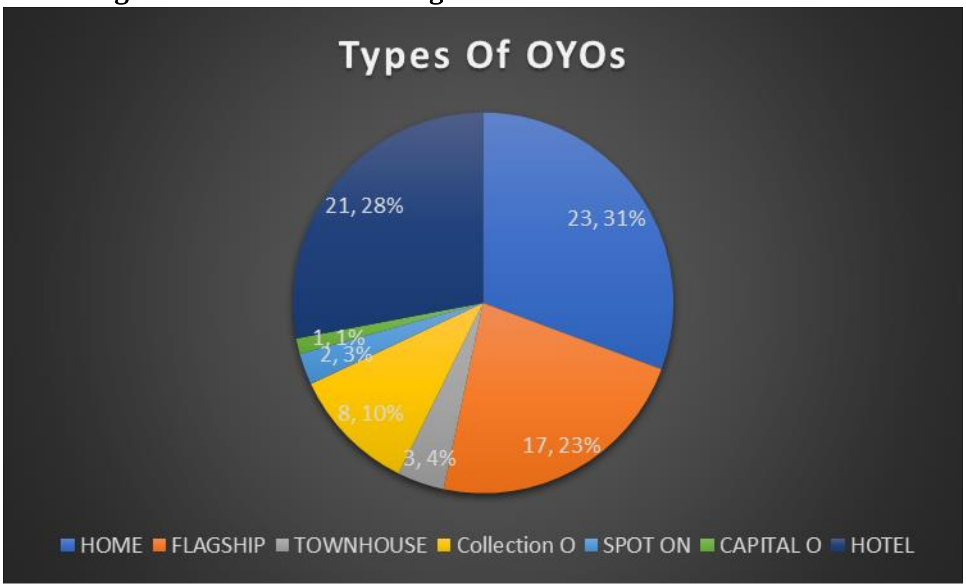

Price
FLAGSHIP
Yes
No
HOME
No
Yes
Conference Room
No
Yes
Bath Tub
No
Yes
Banquet Hall
No
Yes
Twin Single Bed
No
Yes
Queen Sized Bed
No
Yes
Hair Dryer
No
Yes
Dress Hanger
No
Yes
Sofa Beds
No
Yes


Price
Shower Head
Private Entrance
Modern Wardrobe
No
Yes


Rating
BETIM
CALANGUTE
HOTEL
CAPITAL O
SPOT ON
Collection O
No
Yes

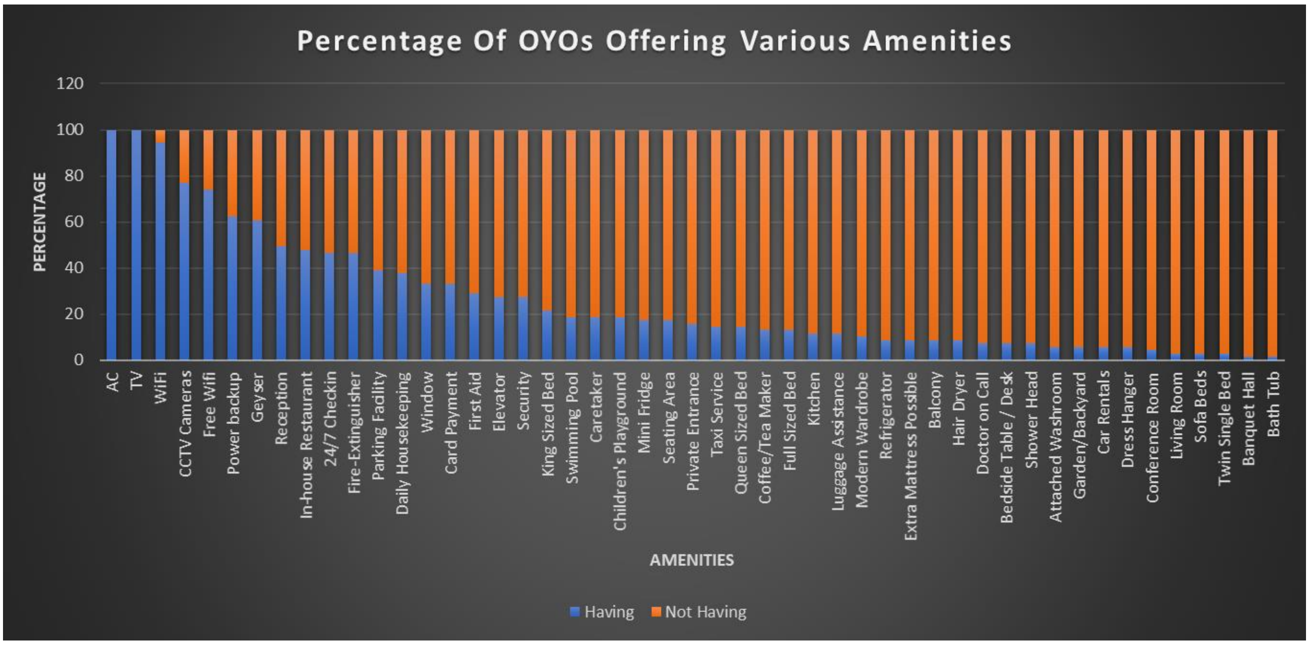


Percentage Of Various OYO Categories In Goa



Top Categories: HOME, HOTEL, FLAGSHIP

Distribution Of Amenities Offered By Various OYOs



Almost all OYOs offer AC,TV and Wi-Fi
More than half offer CCTV cameras, Free Wi-Fi, Power backup, Geyser
Very few OYOs offer Bath Tubs, Banquet Halls, Living Rooms, Conference Rooms

Insights

	% Of hotels Having	correlation with Price
Modern Wardrobe	10.14493	0.362651
Balcony	8.695652	0.303867
Shower Head	7.246377	0.300436
Geyser	60.86957	0.266937
Garden/Backyard	5.797101	0.23174
Private Entrance	15.94203	0.225538
Taxi Service	14.49275	-0.22019
Car Rentals	5.797101	-0.22215

Although the above amenities are more highly correlated to the price as compared to others, we see that most of them are offered by less than 15% of the hotels. Hence OYOs should start offering the above amenities as soon as possible to increase revenue

	% Of hotels Having	correlation with Price
HOME	33.33333	0.17999927
SPOT ON	2.898551	-0.21509843

SPOT ON & Home show the highest correlation with price, however, their presence is less than half. OYO is looked at a more budget hotel, and hence the lower priced categories can fetch better revenue

	% Of hotels Having	correlation with Price
BAGA	4.347826	-0.136747639
PANJIM	14.49275	-0.187245622
CALANGUTE	27.53623	0.134854512

Baga, Panjim & Calangute have the highest correlation with price, OYO Should expand its presence in these 3, with more focus on Panjim

	% Of hotels Having	correlation with ratingRating
HOME	33.33333	-0.211530615
Collection O	11.5942	0.252530611
CAPITAL O	1.449275	-0.253436558

Home, Collection O & Capital O have the best correlation with ratings, OYO should expand its CAPITAL O inorder to gain higher overall ratings

	% Of hotels Having	correlation with ratingRating
CALANGUTE	27.53623	0.267086217
SALIGAO	1.449275	0.210105457

OYOs in Saligao & Calangute have a stronger correlation with rating, however, the number of OYOs in Saligao are very few, hence opening an OYO in saligao can ensure a higher rating

	% Of hotels Having	correlation with ratingRating
Car Rentals	5.797101	0.23817582
Reception	49.27536	0.23667087
Balcony	8.695652	-0.2173356
Dress Hanger	5.797101	-0.2201387
Twin Single Bed	2.898551	-0.2207387
Geyser	60.86957	-0.2464746
Refrigerator	8.695652	-0.2730407
Children's Playground	18.84058	-0.2938626
In-house Restaurant	47.82609	-0.2985888
Shower Head	7.246377	-0.3045747
Bedside Table / Desk	7.246377	-0.3099168
Modern Wardrobe	10.14493	-0.3337285

Although the above amenities especially Modern Wardrobes, Bedside tables, shower heads, restaurants, playgrounds can ensure a higher rating, less than 20% of hotels offer these services

Results from SVM
Prediction of rating of new & unrated OYOs

OYO Name	Predicted Rating
POP 85139 Kabir Guest House	4.5
Townhouse OAK 7 Spices	4.5
OYO 83865 Baba Guest House	4.3
OYO 85326 Corinthia Boutique Rooms	4.3
OYO 85229 Trivikram Krupa Guest House	4.2
OYO 85106 Hotel Elvin,s Place	4.2
OYO 84752 Hotel Joy Guest House	4.2
OYO 84895 Hotel Maples Calangute Pristne	4.2
OYO 83955 Hotel Log Inn	3.5
OYO 84393 David Holiday Home	2.9