In [ ]:

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.preprocessing import OrdinalEncoder
from statsmodels.formula.api import ols
from statsmodels.api import qqplot
from itertools import product
path='/Users/herlihpj/Desktop/Data Analytics/D214 - Capstone/'

sd_airbnb=pd.read_csv(path+'AirBNB Prepared.csv', index_col=1)
print(sd_airbnb.shape)
print(sd_airbnb.columns)
print(sd_airbnb.info())
print(sd_airbnb.describe())

print('======================','Prediction Model','======================')

#Function that builds model formula
def twoway_formula_builder(target_variable, predictor_variables):
    formula=target_variable+' ~ ('
    end=len(predictor_variables)-1
    #use enumerate to get first and last for string manipulation
    for count, value in enumerate(predictor_variables):
        if count==end:
            formula=formula+value
        else:
            formula=formula+value+' + '
    formula=formula+') ** 2 + 0'
    return formula

predictors=['room_type', 'accommodates', 'bathrooms', 'bedrooms', 'beds','host_is_superhost','cleaning_fee','instant_bool
target='nightly_price'
print(twoway_formula_builder(target,predictors))
#Initial Model
i_mdl_airbnbs=ols(twoway_formula_builder(target,predictors),data=sd_airbnb).fit()
print(i_mdl_airbnbs.summary())

print('Final Model')
#Reduced model:
```

```
mdl_airbnb=ols("nightly_price ~ accommodates * bathrooms * room_type + 0",data=sd_airbnb).fit()
#returns intercept and slope?
print('Parameters: ')
print(mdl_airbnb.params)
print('Summary: ')
print(mdl_airbnb.summary())
all_resid=mdl_airbnb.resid
print('Sum of Residuals', all_resid.sum()) #Sum of residuals should be ~0

#Prediction Values
#Create exploratory data for the model
print('Prediction Values: ')
accommodates = np.arange(1, 18, 1)
baths=np.arange(0, 10, 1)
room_type=sd_airbnb['room_type'].unique()
#creates an array of all possible number of guests and number of bathrooms, and roomtype combinations
p=product(accommodates,baths,room_type)
print(p)
#Create a dataframe with all the exploratory values
expl_data = pd.DataFrame(p,columns=["accommodates","bathrooms","room_type"])
#Adds the predicted values to the new data frame
pred_data = expl_data.assign(nightly_prices = mdl_airbnb.predict(expl_data))
print(pred_data)

#Scatter plot of the data
sns.scatterplot(x='accommodates',y="nightly_price",data=sd_airbnb, hue='room_type')#'bedrooms', 'bathrooms'
plt.show()
#Visualize how the predicted values fit the data (Waskom)
sns.scatterplot(x='accommodates',y="nightly_price",data=sd_airbnb, hue='room_type')#'bedrooms', 'bathrooms'
sns.regplot(x='accommodates',y="nightly_prices", data=pred_data, ci=None, line_kws={"color": "black"}, scatter=False)
sns.regplot(x='accommodates',y="nightly_prices", data=pred_data[pred_data.room_type=='Entire home/apt'],
            ci=None, line_kws={"color": "orange"}, scatter=False)
sns.regplot(x='accommodates',y="nightly_prices", data=pred_data[pred_data.room_type=='Private room'],
            ci=None, line_kws={"color": "blue"}, scatter=False)
sns.regplot(x='accommodates',y="nightly_prices", data=pred_data[pred_data.room_type=='Shared room'],
            ci=None, line_kws={"color": "green"}, scatter=False)
plt.show()

#Residual Plots

#Check the model
residuals=mdl_addcharge.resid
R_squared=residuals**2
#calc RSE
mse=mdl_addcharge.mse_resid
```

```
rse=np.sqrt(mse)
print('RSE: ', rse)

#Residual Plots
#entire home
sns.residplot(x="accommodates", y="nightly_price", data=sd_airbnb[sd_airbnb.room_type=='Entire home/apt'], lowess=True)
plt.title('Room Type=Entire Home/Apt Residual Plot')
plt.xlabel("Fitted values")
plt.ylabel("Residuals")
plt.show()
#private room
sns.residplot(x="accommodates", y="nightly_price", data=sd_airbnb[sd_airbnb.room_type=='Private room'], lowess=True)
plt.title('Room Type: Private Room Residual Plot')
plt.xlabel("Fitted values")
plt.ylabel("Residuals")
plt.show()

#Q-Q plot
qqplot(data=mdl_airbnb.resid, fit=True, line="45")
plt.show()

pred_data.to_csv(path+'Airbnb Predicted Prices.csv')
```

```
(10375, 28)
Index(['index', 'host_name', 'host_since', 'host_is_superhost', 'neighborhood',
       'city', 'state', 'latitude', 'longitude', 'is_location_exact',
       'property_type', 'room_type', 'accommodates', 'bathrooms', 'bedrooms',
       'beds', 'amenities', 'nightly_price', 'security_deposit',
       'cleaning_fee', 'guests_included', 'extra_people', 'number_of_reviews',
       'number_of_stays', 'first_review', 'last_review', 'instant_bookable',
       'cancellation_policy'],
      dtype='object')
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10375 entries, 1756516 to 7192087
Data columns (total 28 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   index                10375 non-null  int64  
 1   host_name            10375 non-null  object 
 2   host_since           10375 non-null  object 
 3   host_is_superhost    10375 non-null  object 
 4   neighborhood         10375 non-null  object 
 5   city                 10375 non-null  object 
 6   state                10375 non-null  object 
 7   latitude             10375 non-null  float64
 8   longitude            10375 non-null  float64
 9   is_location_exact    10375 non-null  object 
 10  property_type        10375 non-null  object 
 11  room_type            10375 non-null  object 
 12  accommodates         10375 non-null  int64  
 13  bathrooms            10375 non-null  float64
 14  bedrooms             10375 non-null  int64  
 15  beds                 10375 non-null  int64  
 16  amenities            10375 non-null  object 
 17  nightly_price        10375 non-null  float64
 18  security_deposit     10375 non-null  float64
 19  cleaning_fee         10375 non-null  float64
 20  guests_included      10375 non-null  int64  
 21  extra_people         10375 non-null  float64
 22  number_of_reviews    10375 non-null  int64  
 23  number_of_stays      10375 non-null  int64  
 24  first_review         8930 non-null   object 
 25  last_review          8930 non-null   object 
 26  instant_bookable     10375 non-null  object 
 27  cancellation_policy  10375 non-null  object 
dtypes: float64(7), int64(7), object(14)
memory usage: 2.3+ MB
None
```

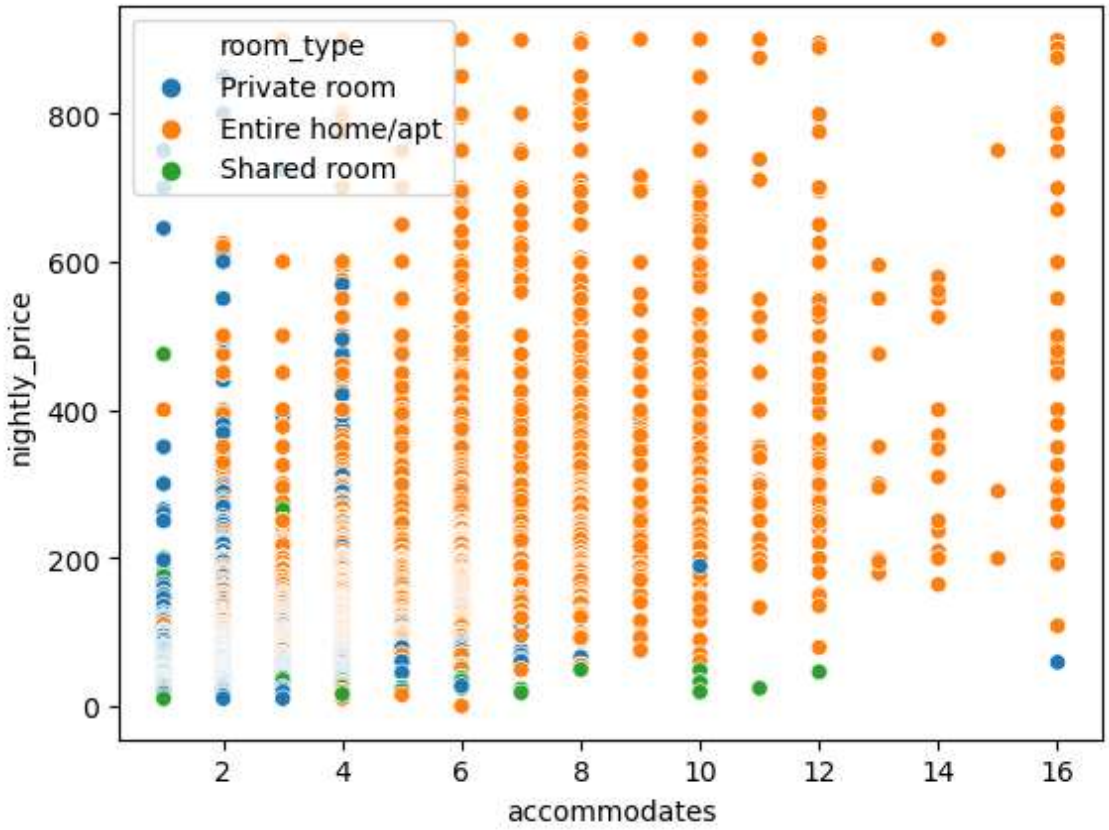
```
              index      latitude     longitude  accommodates     bathrooms  \
count  10375.000000  10375.000000  10375.000000  10375.000000  10375.000000
mean    6167.228627     32.771146   -117.180345      4.030940      1.356289
std     3595.709408      0.059186      0.058266      2.509497      0.652901
min        4.000000     32.662760   -117.281750      1.000000      0.000000
25%     3050.500000     32.727275   -117.240400      2.000000      1.000000
50%     6204.000000     32.756040   -117.165280      4.000000      1.000000
75%     9246.500000     32.797185   -117.140385      6.000000      2.000000
max    13042.000000     33.086070   -117.006610     16.000000      8.000000

           bedrooms          beds  nightly_price  security_deposit  \
count  10375.000000  10375.000000   10375.000000      10375.000000
mean       1.441253      2.117398     168.821108        181.408964
std        0.971343      1.533540     133.612435        236.543889
min        0.000000      0.000000       0.000000          0.000000
25%        1.000000      1.000000      79.000000          0.000000
50%        1.000000      2.000000     128.000000        100.000000
75%        2.000000      3.000000     208.500000        300.000000
max       10.000000     22.000000     900.000000       1004.000000

       cleaning_fee  guests_included  extra_people  number_of_reviews  \
count  10375.000000     10375.000000  10375.000000       10375.000000
mean      71.275759         1.991711     10.373590          38.647325
std       63.865850         1.784821     14.308702          63.048870
min        0.000000         1.000000      0.000000           0.000000
25%       20.000000         1.000000      0.000000           2.000000
50%       55.000000         1.000000      0.000000          13.000000
75%      100.000000         2.000000     20.000000          48.000000
max      265.000000        16.000000     65.000000         642.000000

       number_of_stays
count     10375.000000
mean         77.294651
std         126.097741
min           0.000000
25%           4.000000
50%          26.000000
75%          96.000000
max        1284.000000
======================== Prediction Model ========================
nightly_price ~ (room_type + accommodates + bathrooms + bedrooms + beds + host_is_superhost + cleaning_fee + instant_boo
kable) ** 2 + 0
                            OLS Regression Results
==============================================================================
Dep. Variable:          nightly_price   R-squared:                       0.499
```

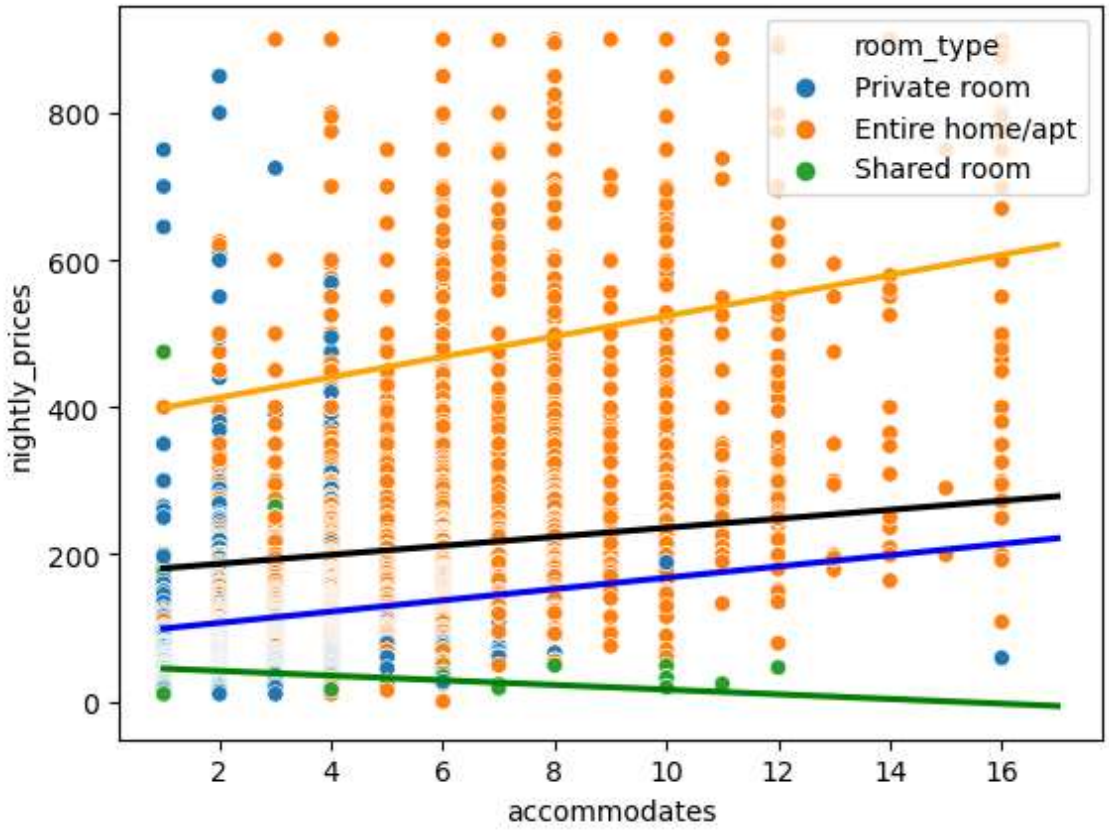
Model:	OLS	Adj. R-squared:	0.497
Method:	Least Squares	F-statistic:	239.7
Date:	Fri, 10 Feb 2023	Prob (F-statistic):	0.00
Time:	09:42:19	Log-Likelihood:	-61916.
No. Observations:	10375	AIC:	1.239e+05
Df Residuals:	10331	BIC:	1.242e+05
Df Model:	43		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
room_type[Entire home/apt]	46.7426	7.647	6.113	0.000	31.753	61.732
room_type[Private room]	56.7697	7.817	7.262	0.000	41.446	72.093
room_type[Shared room]	21.2565	8.995	2.363	0.018	3.625	38.888
host_is_superhost[T.t]	-27.7751	5.966	-4.656	0.000	-39.470	-16.081
instant_bookable[T.t]	-5.4986	5.790	-0.950	0.342	-16.848	5.851
room_type[T.Private room]:host_is_superhost[T.t]	14.8375	5.366	2.765	0.006	4.320	25.355
room_type[T.Shared room]:host_is_superhost[T.t]	60.4417	19.568	3.089	0.002	22.084	98.799
room_type[T.Private room]:instant_bookable[T.t]	8.5615	5.146	1.664	0.096	-1.526	18.649
room_type[T.Shared room]:instant_bookable[T.t]	34.8547	18.496	1.884	0.060	-1.402	71.111
host_is_superhost[T.t]:instant_bookable[T.t]	-0.6108	3.932	-0.155	0.877	-8.318	7.097
accommodates	26.8877	2.509	10.716	0.000	21.969	31.806
room_type[T.Private room]:accommodates	-1.0044	2.705	-0.371	0.710	-6.306	4.297
room_type[T.Shared room]:accommodates	-20.2287	5.280	-3.831	0.000	-30.579	-9.878
accommodates:host_is_superhost[T.t]	-7.0474	1.811	-3.891	0.000	-10.597	-3.497
accommodates:instant_bookable[T.t]	-0.1239	1.739	-0.071	0.943	-3.532	3.284
bathrooms	34.4314	7.566	4.551	0.000	19.601	49.262
room_type[T.Private room]:bathrooms	-39.0080	6.619	-5.893	0.000	-51.983	-26.033
room_type[T.Shared room]:bathrooms	-43.0940	8.004	-5.384	0.000	-58.784	-27.404
bathrooms:host_is_superhost[T.t]	-13.5830	4.461	-3.045	0.002	-22.328	-4.838
bathrooms:instant_bookable[T.t]	-1.3966	4.434	-0.315	0.753	-10.089	7.296
bedrooms	-6.2859	4.827	-1.302	0.193	-15.749	3.177
room_type[T.Private room]:bedrooms	-7.1571	5.857	-1.222	0.222	-18.638	4.324
room_type[T.Shared room]:bedrooms	21.2565	8.995	2.363	0.018	3.625	38.888
bedrooms:host_is_superhost[T.t]	-8.0781	3.674	-2.199	0.028	-15.280	-0.877
bedrooms:instant_bookable[T.t]	-1.4713	3.493	-0.421	0.674	-8.317	5.375
beds	-4.8934	3.742	-1.308	0.191	-12.229	2.442
room_type[T.Private room]:beds	4.7136	4.523	1.042	0.297	-4.152	13.579
room_type[T.Shared room]:beds	4.8943	6.612	0.740	0.459	-8.066	17.855
beds:host_is_superhost[T.t]	13.8998	2.539	5.474	0.000	8.923	18.877
beds:instant_bookable[T.t]	-3.3216	2.514	-1.321	0.186	-8.250	1.607
cleaning_fee	-0.0958	0.062	-1.548	0.122	-0.217	0.026
room_type[T.Private room]:cleaning_fee	-0.1753	0.076	-2.312	0.021	-0.324	-0.027
room_type[T.Shared room]:cleaning_fee	-0.0512	0.481	-0.106	0.915	-0.995	0.892
host_is_superhost[T.t]:cleaning_fee	0.4668	0.046	10.124	0.000	0.376	0.557

```
cleaning_fee:instant_bookable[T.t]                  0.1623      0.041      3.962      0.000       0.082       0.243
accommodates:bathrooms                              1.1449      1.357      0.844      0.399      -1.515       3.805
accommodates:bedrooms                              -5.2597      0.934     -5.633      0.000      -7.090      -3.429
accommodates:beds                                  -0.2764      0.328     -0.842      0.400      -0.920       0.367
accommodates:cleaning_fee                          -0.0382      0.014     -2.769      0.006      -0.065      -0.011
bathrooms:bedrooms                                 13.8011      2.732      5.052      0.000       8.446      19.156
bathrooms:beds                                     -4.0139      1.779     -2.256      0.024      -7.501      -0.526
bathrooms:cleaning_fee                             -0.0113      0.039     -0.288      0.773      -0.088       0.066
bedrooms:beds                                       4.4498      1.123      3.963      0.000       2.249       6.651
bedrooms:cleaning_fee                               0.1976      0.032      6.231      0.000       0.135       0.260
beds:cleaning_fee                                   0.0152      0.019      0.812      0.417      -0.021       0.052
==============================================================================
Omnibus:                     5068.247   Durbin-Watson:                   1.240
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            46417.889
Skew:                           2.146   Prob(JB):                         0.00
Kurtosis:                      12.431   Cond. No.                     1.35e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 3.79e-23. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
Final Model
Parameters:
room_type[Entire home/apt]                          1.221138
room_type[Private room]                            30.807801
room_type[Shared room]                             61.536756
accommodates                                       18.497899
accommodates:room_type[T.Private room]              4.982033
accommodates:room_type[T.Shared room]             -18.874343
bathrooms                                          85.354119
bathrooms:room_type[T.Private room]               -71.893911
bathrooms:room_type[T.Shared room]                -88.411216
accommodates:bathrooms                             -1.029512
accommodates:bathrooms:room_type[T.Private room]   -2.483620
accommodates:bathrooms:room_type[T.Shared room]     0.406653
dtype: float64
Summary:
                            OLS Regression Results
==============================================================================
Dep. Variable:          nightly_price   R-squared:                       0.464
Model:                            OLS   Adj. R-squared:                  0.464
Method:                 Least Squares   F-statistic:                     816.2
Date:                Fri, 10 Feb 2023   Prob (F-statistic):               0.00
Time:                        09:42:20   Log-Likelihood:                -62269.
```

```
No. Observations:               10375   AIC:                          1.246e+05
Df Residuals:                   10363   BIC:                          1.246e+05
Df Model:                          11
Covariance Type:            nonrobust
=================================================================================================================
                                                    coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------------------------
room_type[Entire home/apt]                        1.2211      6.221      0.196      0.844     -10.973      13.416
room_type[Private room]                          30.8078      9.661      3.189      0.001      11.870      49.746
room_type[Shared room]                           61.5368     18.397      3.345      0.001      25.475      97.598
accommodates                                     18.4979      1.165     15.880      0.000      16.215      20.781
accommodates:room_type[T.Private room]            4.9820      3.845      1.296      0.195      -2.554      12.518
accommodates:room_type[T.Shared room]           -18.8743      7.693     -2.454      0.014     -33.953      -3.795
bathrooms                                        85.3541      4.492     19.000      0.000      76.548      94.160
bathrooms:room_type[T.Private room]             -71.8939      8.211     -8.756      0.000     -87.989     -55.798
bathrooms:room_type[T.Shared room]              -88.4112      8.207    -10.772      0.000    -104.499     -72.323
accommodates:bathrooms                           -1.0295      0.593     -1.737      0.082      -2.191       0.132
accommodates:bathrooms:room_type[T.Private room] -2.4836      2.159     -1.151      0.250      -6.715       1.748
accommodates:bathrooms:room_type[T.Shared room]   0.4067      3.766      0.108      0.914      -6.976       7.789
==============================================================================
Omnibus:                     4787.832   Durbin-Watson:                   1.154
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            37309.940
Skew:                           2.056   Prob(JB):                         0.00
Kurtosis:                      11.331   Cond. No.                         228.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Sum of Residuals 1.1092197382822633e-08
Prediction Values:
<itertools.product object at 0x000002034E8713C0>
     accommodates  bathrooms        room_type  nightly_prices
0               1          0     Private room       54.287732
1               1          0  Entire home/apt       19.719036
2               1          0      Shared room       61.160311
3               1          1     Private room       64.234808
4               1          1  Entire home/apt      104.043643
..            ...        ...              ...             ...
505            17          8  Entire home/apt      858.504735
506            17          8      Shared room      -54.028473
507            17          9     Private room       13.599370
508            17          9  Entire home/apt      926.357150
509            17          9      Shared room      -67.674182

[510 rows x 4 columns]
```


room_type
Private room
Entire home/apt
Shared room
nightly_price
0
200
400
600
800
2
4
6
8
10
12
14
16
accommodates




```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_7764\757019946.py in <module>
     81
     82 #Check the model
---> 83 residuals=mdl_addcharge.resid
     84 R_squared=residuals**2
     85 #calc RSE

NameError: name 'mdl_addcharge' is not defined
```

In [ ]: