Jumana Rahman

Types of Hotels Booked based on Travel Destination and Customer Necessities

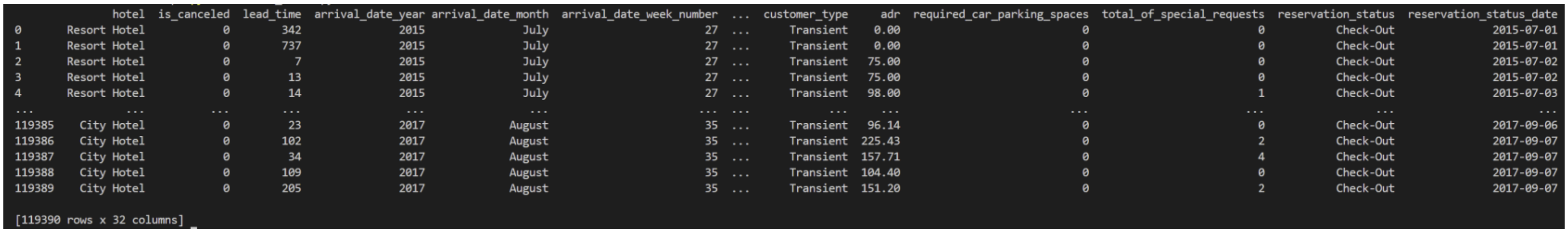
Abstract

When booking flights and hotels for vacations, one may wonder what type of hotel to book, what type of hotels provides better services for traveling with a car, whether that hotel is children friendly, etc. Although these things are naturally considered when booking hotels for vacations, such aspects can assist in picking out what types of hotels would be best for a specific country or for a specific time of year. Further analysis can also be done for the city and resort hotels to see patterns for what time of year people are most likely to book stays in and what time of year customers are most likely to bring children along. By conducting such studies and analyzing patterns in a dataset consisting of hotel booking information, it can help hotels provide services to their customer's based on the trends and patterns that may be visible along with help people in deciding on what type of hotel to book for their vacations. This is a study conducted on a dataset containing information on hotel bookings in order to find patterns and predict what type of hotels people prefer to book for vacations based on several aspects like the country, time of year, number of adults, number of children, week nights and weekend nights stayed in, average daily rate, and required parking spaces.

Collect and organize the data

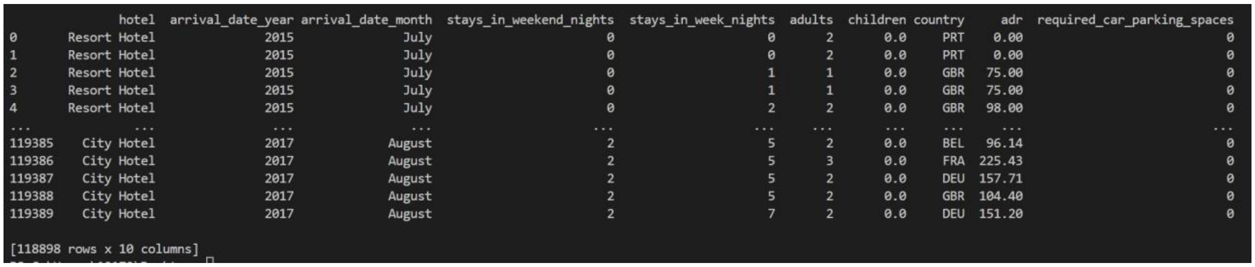
After loading the hotel_bookings.csv file into a dataframe and printing it, I saw that there was 119390 rows and 32 columns, which meant that a total of 119390 bookings were in this dataset. However, I did not need all 32 columns so I dropped the columns that were irrelevant to the study and made two data frames. hotelData will be used to create a model and make predictions and the other data frame, hotelBookingsData, will be used to analyze any patterns, connections, or relations within the dataset. Amongst the hotelBookingsData, two more dataframes will be creates, cityHotelBookings and resortHotelBookings in order to observe the patterns for the bookings for city hotels and resort hotels.

Original Data frame of Dataset:

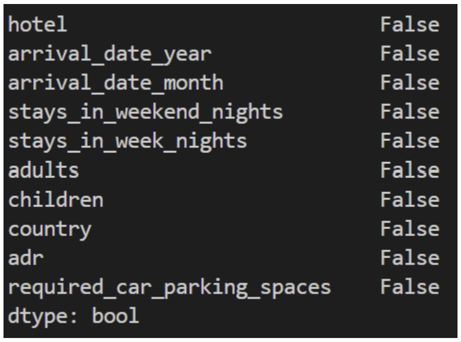


```
               hotel  is_canceled  lead_time  arrival_date_year arrival_date_month  arrival_date_week_number  ...  customer_type     adr  required_car_parking_spaces  total_of_special_requests  reservation_status  reservation_status_date
0       Resort Hotel            0        342               2015               July                        27  ...      Transient    0.00                            0                          0           Check-Out               2015-07-01
1       Resort Hotel            0        737               2015               July                        27  ...      Transient    0.00                            0                          0           Check-Out               2015-07-01
2       Resort Hotel            0          7               2015               July                        27  ...      Transient   75.00                            0                          0           Check-Out               2015-07-02
3       Resort Hotel            0         13               2015               July                        27  ...      Transient   75.00                            0                          0           Check-Out               2015-07-02
4       Resort Hotel            0         14               2015               July                        27  ...      Transient   98.00                            0                          1           Check-Out               2015-07-03
...              ...          ...        ...                ...                ...                       ...  ...            ...     ...                          ...                        ...                 ...                      ...
119385    City Hotel            0         23               2017             August                        35  ...      Transient   96.14                            0                          0           Check-Out               2017-09-06
119386    City Hotel            0        102               2017             August                        35  ...      Transient  225.43                            0                          2           Check-Out               2017-09-07
119387    City Hotel            0         34               2017             August                        35  ...      Transient  157.71                            0                          4           Check-Out               2017-09-07
119388    City Hotel            0        109               2017             August                        35  ...      Transient  104.40                            0                          0           Check-Out               2017-09-07
119389    City Hotel            0        205               2017             August                        35  ...      Transient  151.20                            0                          2           Check-Out               2017-09-07

[119390 rows x 32 columns]
```

hotelBookingsData after dropping all null values:

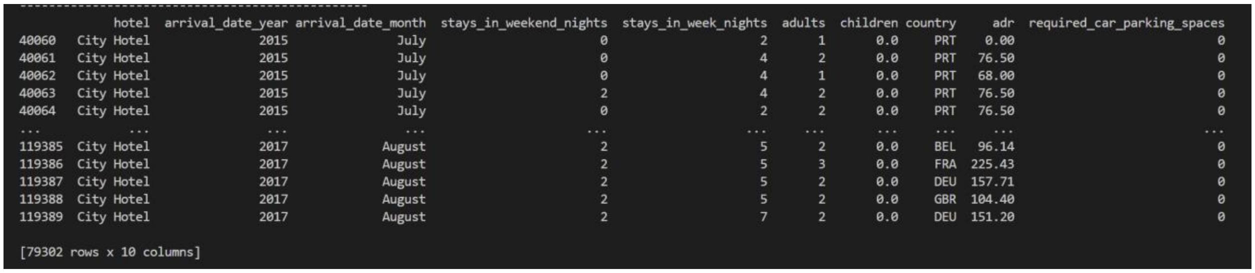


```
               hotel  arrival_date_year arrival_date_month  stays_in_weekend_nights  stays_in_week_nights  adults  children country     adr  required_car_parking_spaces
0       Resort Hotel               2015               July                        0                     0       2       0.0     PRT    0.00                            0
1       Resort Hotel               2015               July                        0                     0       2       0.0     PRT    0.00                            0
2       Resort Hotel               2015               July                        0                     1       1       0.0     GBR   75.00                            0
3       Resort Hotel               2015               July                        0                     1       1       0.0     GBR   75.00                            0
4       Resort Hotel               2015               July                        0                     2       2       0.0     GBR   98.00                            0
...              ...                ...                ...                      ...                   ...     ...       ...     ...     ...                          ...
119385    City Hotel               2017             August                        2                     5       2       0.0     BEL   96.14                            0
119386    City Hotel               2017             August                        2                     5       3       0.0     FRA  225.43                            0
119387    City Hotel               2017             August                        2                     5       2       0.0     DEU  157.71                            0
119388    City Hotel               2017             August                        2                     5       2       0.0     GBR  104.40                            0
119389    City Hotel               2017             August                        2                     7       2       0.0     DEU  151.20                            0

[118898 rows x 10 columns]
```



```
hotel                          False
arrival_date_year              False
arrival_date_month             False
stays_in_weekend_nights        False
stays_in_week_nights           False
adults                         False
children                       False
country                        False
adr                            False
required_car_parking_spaces    False
dtype: bool
```

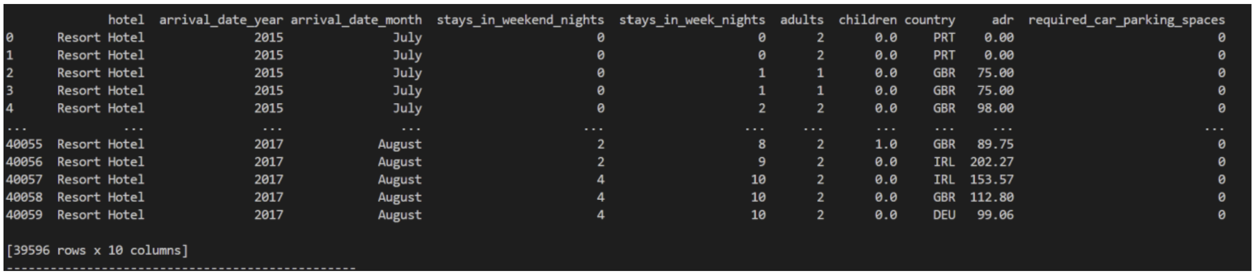
cityHotelBookings:



```
----------------------------------------------
             hotel  arrival_date_year arrival_date_month  stays_in_weekend_nights  stays_in_week_nights  adults  children country     adr  required_car_parking_spaces
40060   City Hotel               2015               July                        0                     2       1       0.0     PRT    0.00                            0
40061   City Hotel               2015               July                        0                     4       2       0.0     PRT   76.50                            0
40062   City Hotel               2015               July                        0                     4       1       0.0     PRT   68.00                            0
40063   City Hotel               2015               July                        2                     4       2       0.0     PRT   76.50                            0
40064   City Hotel               2015               July                        0                     2       2       0.0     PRT   76.50                            0
...            ...                ...                ...                      ...                   ...     ...       ...     ...     ...                          ...
119385  City Hotel               2017             August                        2                     5       2       0.0     BEL   96.14                            0
119386  City Hotel               2017             August                        2                     5       3       0.0     FRA  225.43                            0
119387  City Hotel               2017             August                        2                     5       2       0.0     DEU  157.71                            0
119388  City Hotel               2017             August                        2                     5       2       0.0     GBR  104.40                            0
119389  City Hotel               2017             August                        2                     7       2       0.0     DEU  151.20                            0

[79302 rows x 10 columns]
```

From this we can see that there are a total of 79, 302 bookings that were for city hotels throughout all countries in the dataset.

resortHotelBookings:



```
              hotel  arrival_date_year arrival_date_month  stays_in_weekend_nights  stays_in_week_nights  adults  children country     adr  required_car_parking_spaces
0      Resort Hotel               2015               July                        0                     0       2       0.0     PRT    0.00                            0
1      Resort Hotel               2015               July                        0                     0       2       0.0     PRT    0.00                            0
2      Resort Hotel               2015               July                        0                     1       1       0.0     GBR   75.00                            0
3      Resort Hotel               2015               July                        0                     1       1       0.0     GBR   75.00                            0
4      Resort Hotel               2015               July                        0                     2       2       0.0     GBR   98.00                            0
...             ...                ...                ...                      ...                   ...     ...       ...     ...     ...                          ...
40055  Resort Hotel               2017             August                        2                     8       2       1.0     GBR   89.75                            0
40056  Resort Hotel               2017             August                        2                     9       2       0.0     IRL  202.27                            0
40057  Resort Hotel               2017             August                        4                    10       2       0.0     IRL  153.57                            0
40058  Resort Hotel               2017             August                        4                    10       2       0.0     GBR  112.80                            0
40059  Resort Hotel               2017             August                        4                    10       2       0.0     DEU   99.06                            0

[39596 rows x 10 columns]
----------------------------------------------
```

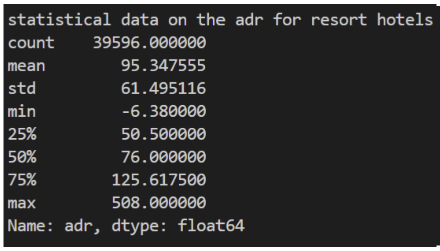
From this we can see that there are a total of 39, 596 bookings that were for resort hotels throughout all countries in the dataset.

Explore the Data Statistically

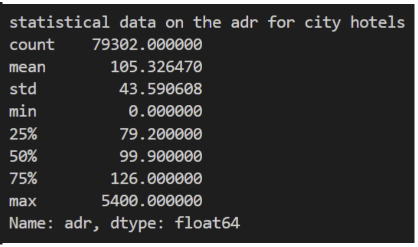
Statistical data for each column

Do a histogram of adr for city hotels and for resort hotels, find the mean value of adr

The adr column in the data frame stands for average daily rate. The value is calculated by diving the sum of all transactions by the total number of staying nights.

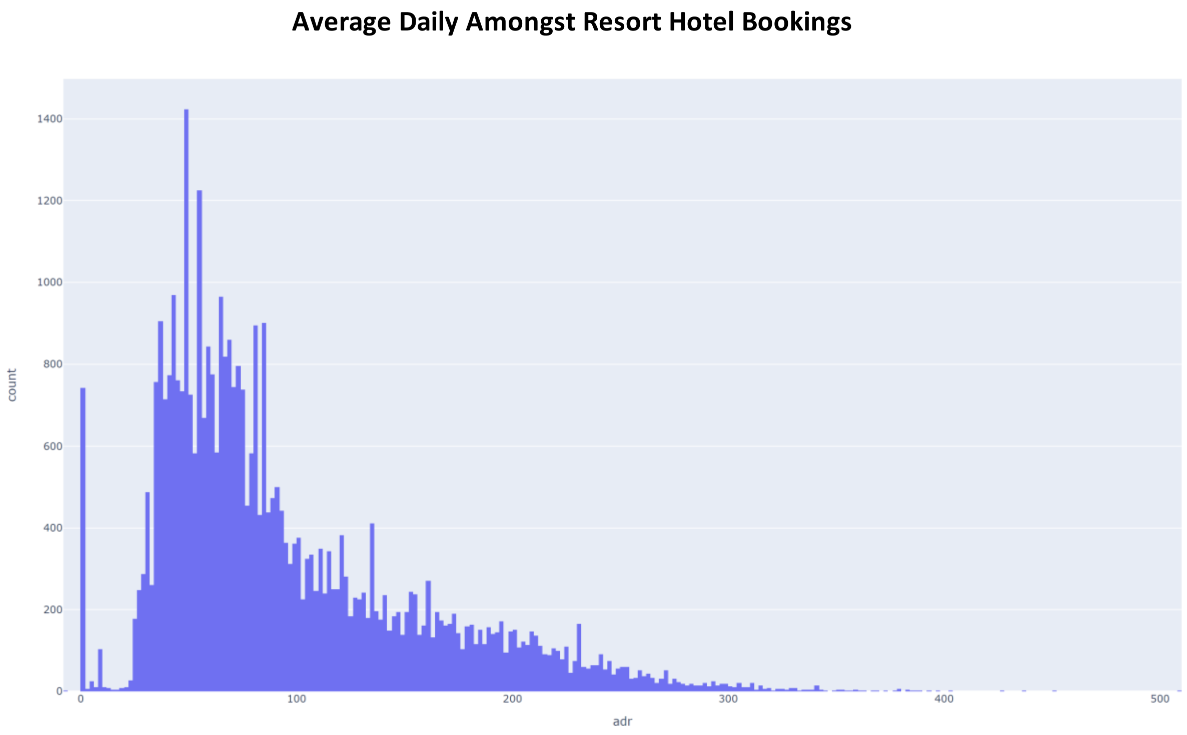


```
statistical data on the adr for resort hotels
count    39596.000000
mean        95.347555
std         61.495116
min         -6.380000
25%         50.500000
50%         76.000000
75%        125.617500
max        508.000000
Name: adr, dtype: float64
```

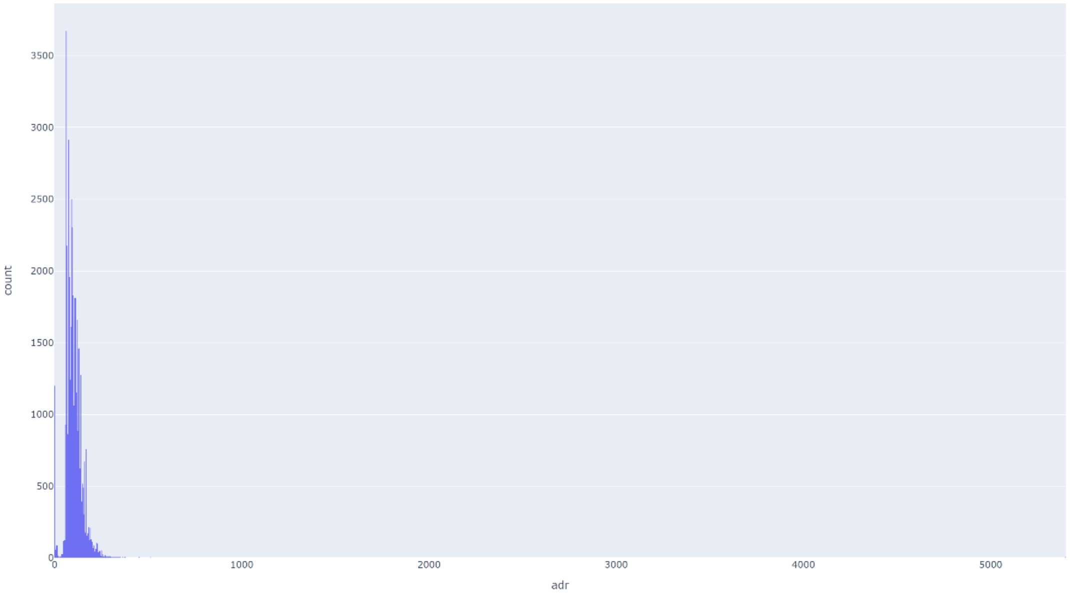


```
statistical data on the adr for city hotels
count    79302.000000
mean       105.326470
std         43.590608
min          0.000000
25%         79.200000
50%         99.900000
75%        126.000000
max       5400.000000
Name: adr, dtype: float64
```

Among the values above, the mean and median for the adr values for city hotels and resort hotels are listed. Both the mean and median values for both resort hotels and city hotels indicate that the histogram will be skewed to the right except that the mean and median for city hotels seem to be closer in value than the mean and median values for the resort hotels.



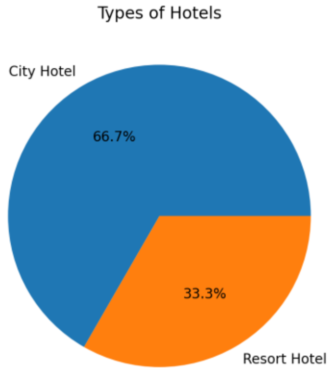
Average Daily Amongst City Hotel Bookings

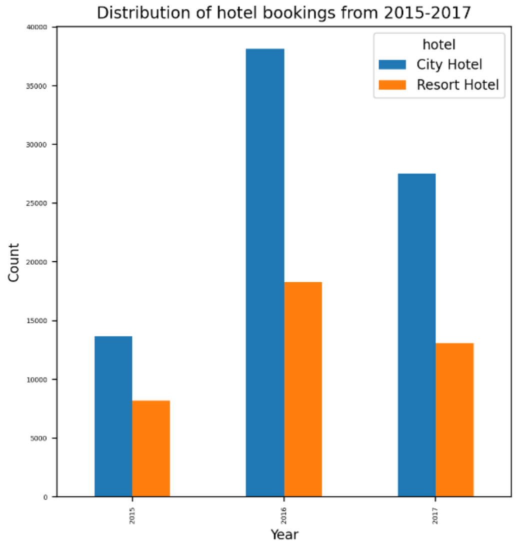


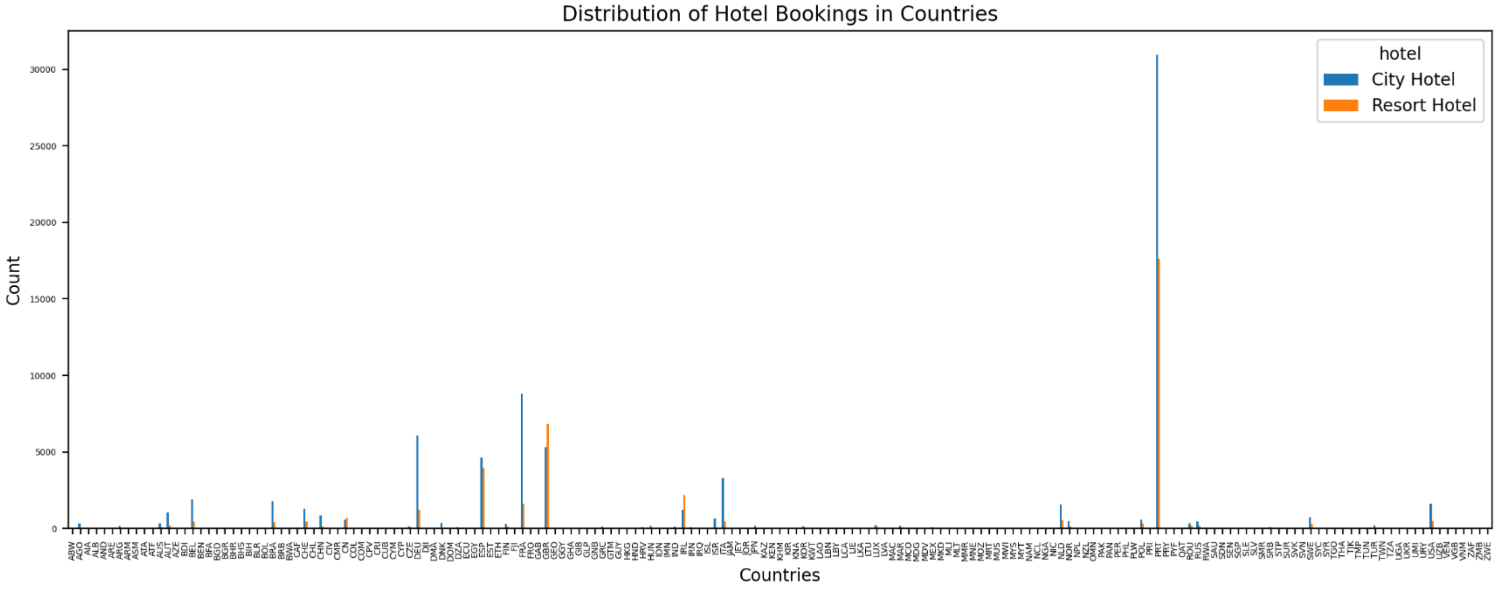
The two graphs above are the histograms showing the distribution of the Average Daily Rates for city and resort hotels and they are both skewed to the right where most of the adr values are around $100. After looking at the statistical data and the histograms, it can be seen that city and resort hotels have around the same average daily rate where the adr value for resort hotel is slightly higher. Both the data for the city hotel and resort hotel have maximum values that serve to be outliers in the dataset where the maximum adr value for resort hotel was $508 and the maximum adr value for city hotel was $5,400. I found this to be surprising because I’ve always thought that resorts and the services available with the resorts are pricier than the rooms and services available at a city hotel.

Connections between columns

Overall distribution of resort hotel bookings and city hotel bookings in the dataset

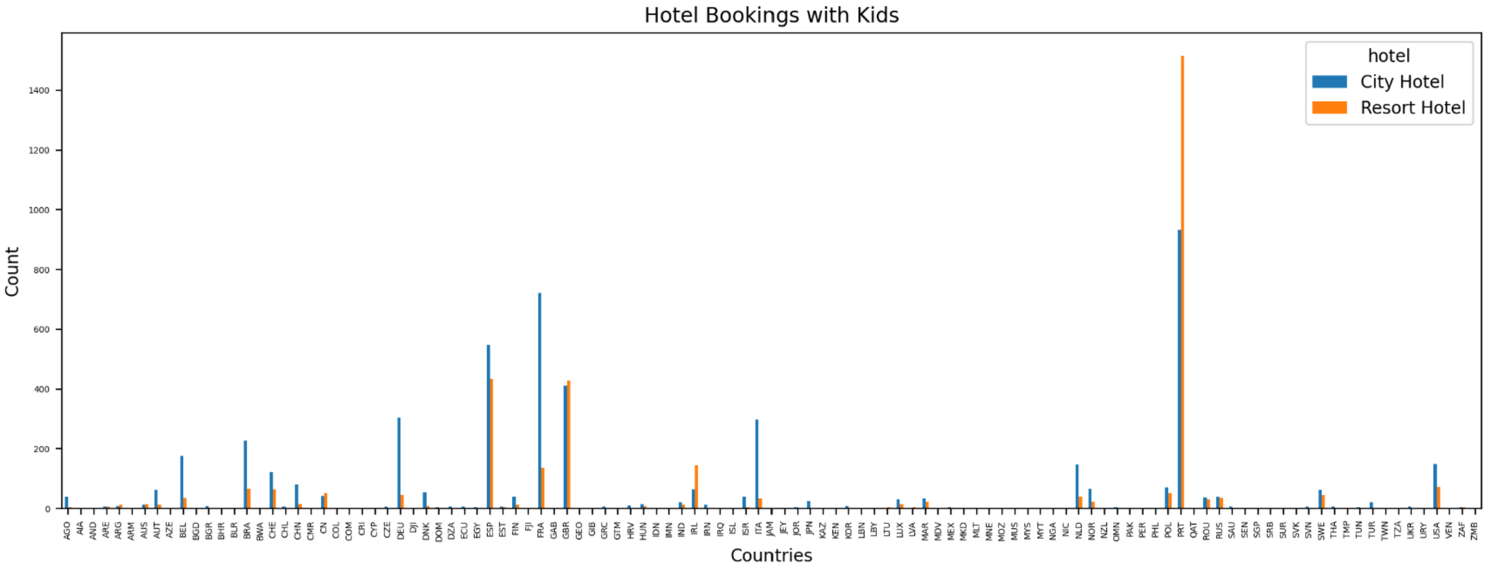






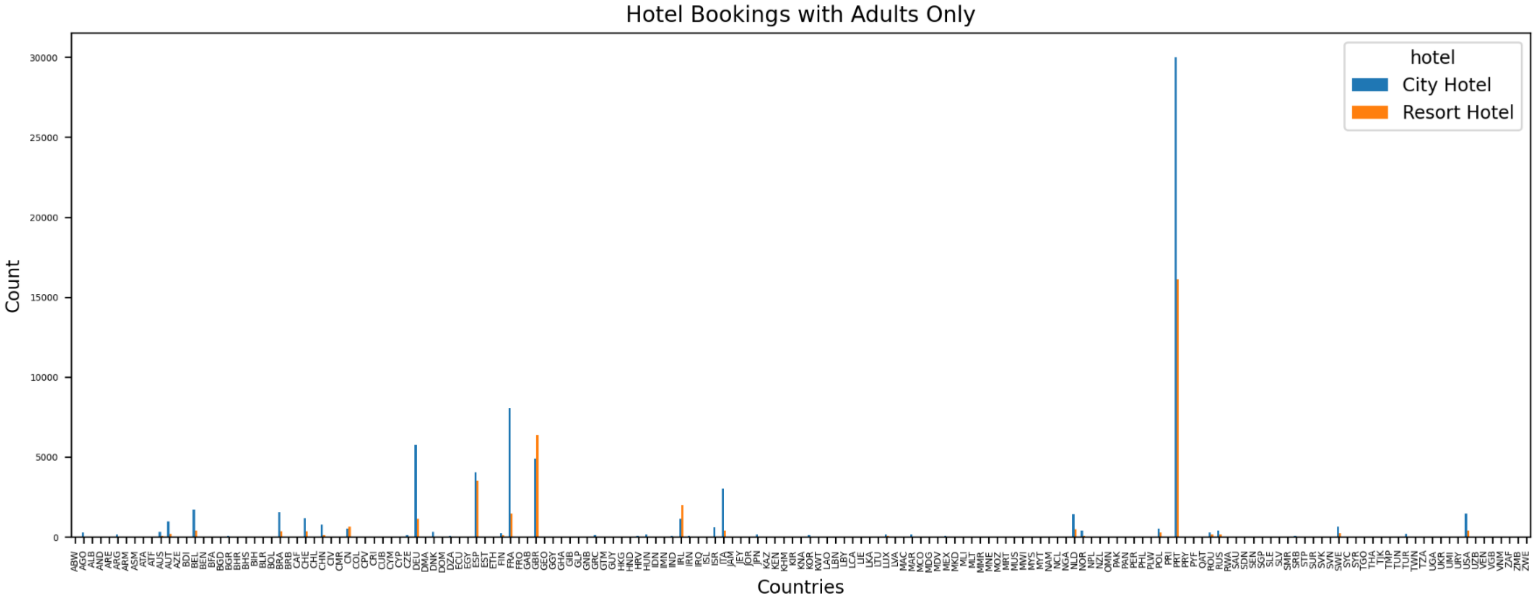
From both the pie chart and the bar graph we can see that city hotel bookings are the more popular choice for type of hotel amongst the bookings present in the dataset. From the pie chart depicting the distribution of hotel type choice, we can see that 66.7% of the entries are city hotel bookings and 33.3% of the entries are resort hotel bookings. In order to get a better view of the distribution of preferred hotel types, a bar graph was constructed to show the choice of hotel type amongst the different countries. After looking at the bar graph, we can see that in most countries, the bookings for city hotels exceed the bookings for resort hotels. However, countries where resort hotels were preferred included the United Kingdom and Ireland. Amongst the countries, Portugal seems to be the most popular for both city hotels and resort hotels. Overall, the top three countries for city hotel bookings are Portugal, France, and Germany and the top three countries for resort hotel bookings are Portugal, United Kingdom, and Spain. Additionally, in the first bar graph, we can see that 2016 was the year that had the most bookings for both city and resort hotels.

Distribution of City and Resort Hotel Bookings with Kids in Different Countries



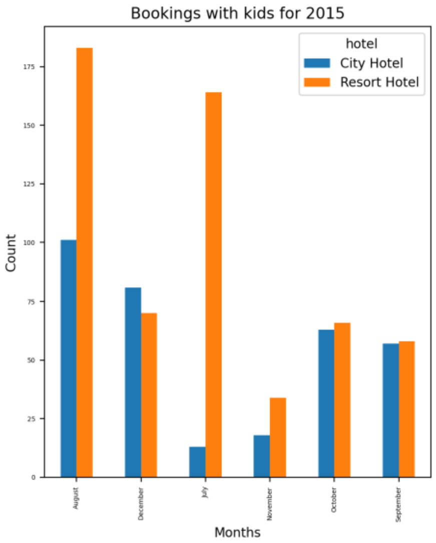
The bar graph above is a visualization of the distribution of hotel bookings including children in different countries. Compared to the previous bar graph, there seems to be more countries where the number of resort hotel bookings exceed the number of city hotel bookings for when children are considered. Now some countries where resorts are preferred are Argentina, Australia, China, United Kingdom, Ireland, and Portugal. The top three countries for resort hotel bookings with kids seem to be Portugal, United Kingdom, and Spain. The top three countries for city hotel bookings with kids seem to be Portugal, France, and Spain.

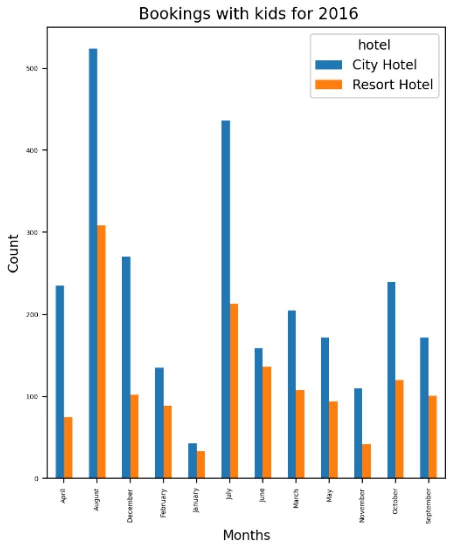
Distribution of City and Resort Hotel Bookings with only adults in Different Countries

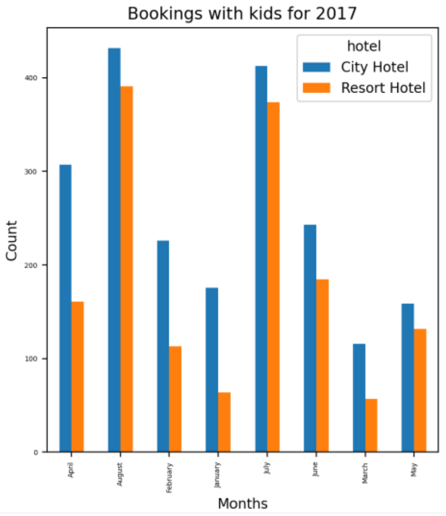


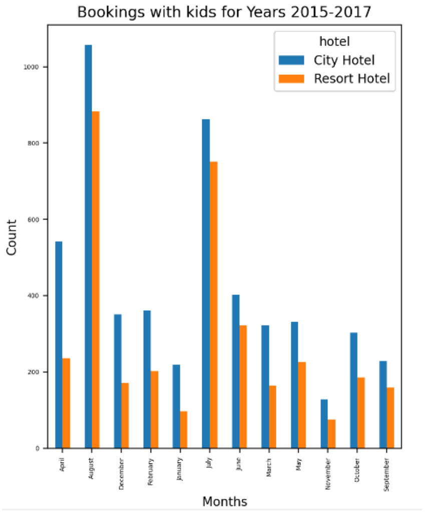
The bar graph above is a visualization of the distribution of hotel bookings that only included adults in different countries. This bar graph is really similar to the bar graph showing the general distribution of the types of hotel bookings in different countries. The top three countries for city hotel and resort hotel bookings remain the same as the bar graph with the general types of hotel distribution amongst the countries in the dataset.

Pattern for Time of year for bookings with children



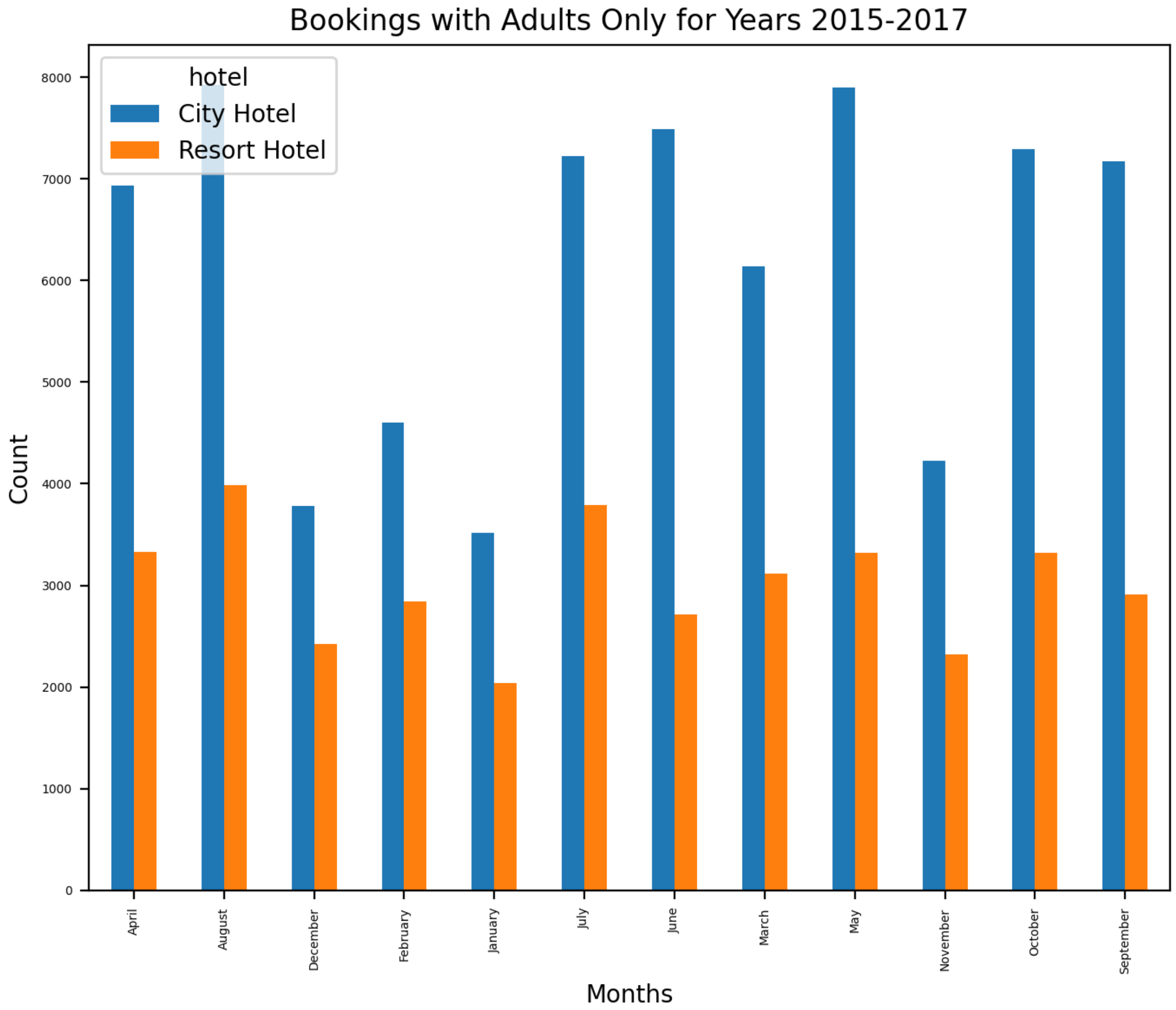






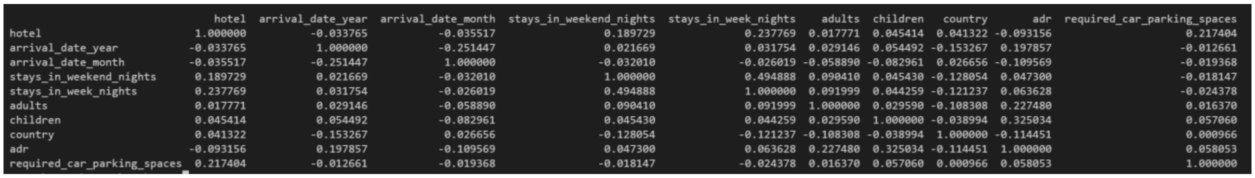
The four bar graphs pictured above show a visualization of the distribution of hotel bookings with kids throughout the months for every year. We can see that resort hotels were most popular in the year 2015, however afterwards, city hotels had started to become the preferred hotel choice. For 2015, the months with the most resort bookings with kids are July and August and the most city hotel bookings with kids are August and December. For 2016 and 2017, the most hotel bookings with kids for both city and resort hotels were August and July. This can be because summer break for children is in July and August and there are more activities to do in the summer at resorts and in the city resulting in more bookings during those months. From this, we can see that both city and resort hotels should prepare for bookings with children during July and August.

Pattern for Time of year for bookings with adults



The bar graph above shows the distribution of hotel bookings amongst bookings with only adults throughout the months. We can see here that the number of bookings for city hotels exceeds the number of bookings for resort hotels every month. The month with the highest number of resort hotel bookings is August and this can be due to August being one of the hottest months of the year. Also, the months with the most city hotel bookings are August and May. Other than that, bookings with only adults also seem to be very common amongst the other months too. This can be because adults that are working have the option to choose when to take off for vacation. From this we can see that most adults tend to book hotels during August and May where most prefer city hotels over resort hotels.

Correlation between columns in the data frame



```
                                hotel  arrival_date_year  arrival_date_month  stays_in_weekend_nights  stays_in_week_nights    adults  children   country       adr  required_car_parking_spaces
hotel                        1.000000          -0.033765           -0.035517                 0.189729              0.237769  0.017771  0.045414  0.041322 -0.093156                     0.217404
arrival_date_year           -0.033765           1.000000           -0.251447                 0.021669              0.031754  0.029146  0.054492 -0.153267  0.197857                    -0.012661
arrival_date_month          -0.035517          -0.251447            1.000000                -0.032010             -0.026019 -0.058890 -0.082961  0.026656 -0.109569                    -0.019368
stays_in_weekend_nights      0.189729           0.021669           -0.032010                 1.000000              0.494888  0.090410  0.045430 -0.128054  0.047300                    -0.018147
stays_in_week_nights         0.237769           0.031754           -0.026019                 0.494888              1.000000  0.091999  0.044259 -0.121237  0.063628                    -0.024378
adults                       0.017771           0.029146           -0.058890                 0.090410              0.091999  1.000000  0.029590 -0.108308  0.227480                     0.016370
children                     0.045414           0.054492           -0.082961                 0.045430              0.044259  0.029590  1.000000 -0.038994  0.325034                     0.057060
country                      0.041322          -0.153267            0.026656                -0.128054             -0.121237 -0.108308 -0.038994  1.000000 -0.114451                     0.000966
adr                         -0.093156           0.197857           -0.109569                 0.047300              0.063628  0.227480  0.325034 -0.114451  1.000000                     0.058053
required_car_parking_spaces  0.217404          -0.012661           -0.019368                -0.018147             -0.024378  0.016370  0.057060  0.000966  0.058053                     1.000000
```

This image above shows the correlation coefficient values between all the columns in the hotelBookingsData data frame.

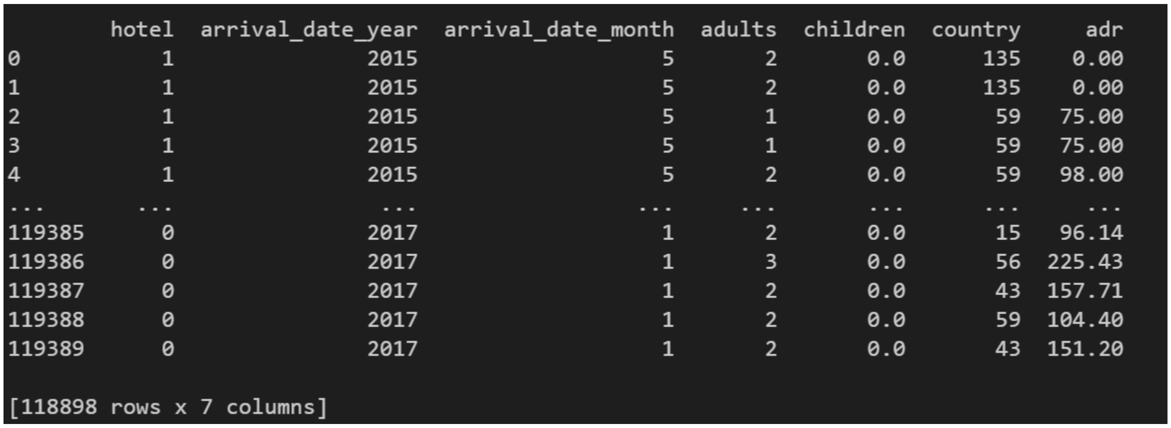
Predict whether a person is more likely to pick city hotel or resort hotel given information about the time of year and country.

Gaussian Model

The Gaussian Model will be used to predict what the next booking will look like and whether based on previous information on bookings, the person will choose a city hotel or resort hotel.

Extract features

The first step was to extract all the necessary features and split the data into train and test data.



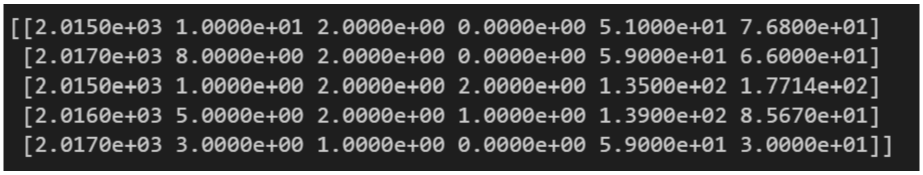
```
        hotel  arrival_date_year  arrival_date_month  adults  children  country     adr
0           1               2015                   5       2       0.0      135    0.00
1           1               2015                   5       2       0.0      135    0.00
2           1               2015                   5       1       0.0       59   75.00
3           1               2015                   5       1       0.0       59   75.00
4           1               2015                   5       2       0.0       59   98.00
...       ...                ...                 ...     ...       ...      ...     ...
119385      0               2017                   1       2       0.0       15   96.14
119386      0               2017                   1       3       0.0       56  225.43
119387      0               2017                   1       2       0.0       43  157.71
119388      0               2017                   1       2       0.0       59  104.40
119389      0               2017                   1       2       0.0       43  151.20

[118898 rows x 7 columns]
```

First, I set the independent variable to be a data frame containing the arrival_date_year, arrival_date_month, adults, children, country, and adr.

Then I set the target variable, y, to the type of hotel chosen.

Train the data so 80% will be training data and 20% will be test data

We get the first five rows of the trained data to be:



```
[[2.0150e+03 1.0000e+01 2.0000e+00 0.0000e+00 5.1000e+01 7.6800e+01]
 [2.0170e+03 8.0000e+00 2.0000e+00 0.0000e+00 5.9000e+01 6.6000e+01]
 [2.0150e+03 1.0000e+00 2.0000e+00 2.0000e+00 1.3500e+02 1.7714e+02]
 [2.0160e+03 5.0000e+00 2.0000e+00 1.0000e+00 1.3900e+02 8.5670e+01]
 [2.0170e+03 3.0000e+00 1.0000e+00 0.0000e+00 5.9000e+01 3.0000e+01]]
```

Standardization

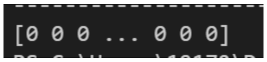
Did sc.fit_transform to our x_train dataframe to fit it over the data and transform it.

After that, did sc.transform(x_test) to perform centering and scaling for the x_test.

Import our Gaussian Model

Now it was time to train the model using the training set using the gnb.fit() function.

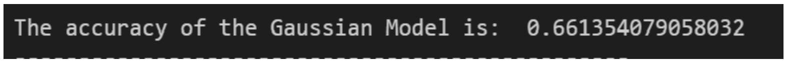
Now to predict the target response for our test dataset and store it in y_predict which looks like:



```
--------------------
[0 0 0 ... 0 0 0]
```

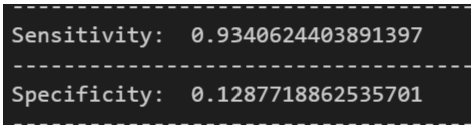
EVALUATE THE MODEL’S OUTPUT

Accuracy measures the correctly classified observations in the model. The program outputted a value of approximately 0.66135 for our Gaussian model which means that the model isn’t the best but may be okay to use.



```
The accuracy of the Gaussian Model is:  0.661354079058032
------------------------------------------------
```

Additionally, we can evaluate the model’s output by looking at the confusion matrix to see the amount of false positives and negatives there are. We can also use this matrix to calculate the sensitivity and specificity values. The sensitivity measures the proportion of positives that are correctly identified in the model and for the model produced, the sensitivity value is approximately 0.934. The specificity value measures the proportion of negatives that are correctly identified in the model. The model produced by this program, we get a specificity value of 0.12877. The value for sensitivitiy is high which indicates a good model, however due to the specificity value being low, it may lead to this model being an okay model but not the absolute best to use.

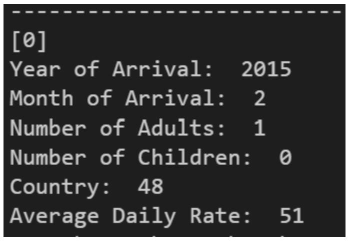


```
-----------------------------------
Sensitivity:  0.9340624403891397
-----------------------------------
Specificity:  0.1287718862535701
-----------------------------------
```

Predicted Output from the Gaussian Model

Result: [0]

The Gaussian model seemed to be an okay enough model to use to predict what type of hotel the next person to book a hotel would choose. In order to do so, I used the gnb.predict function and got a result of 0 which is city hotel. This seemed to be a reasonable prediction because more than half of the original dataset had city hotels bookings. More prediction were conducted on the dataset to determine how the student spends their day and the results we got is as follows:

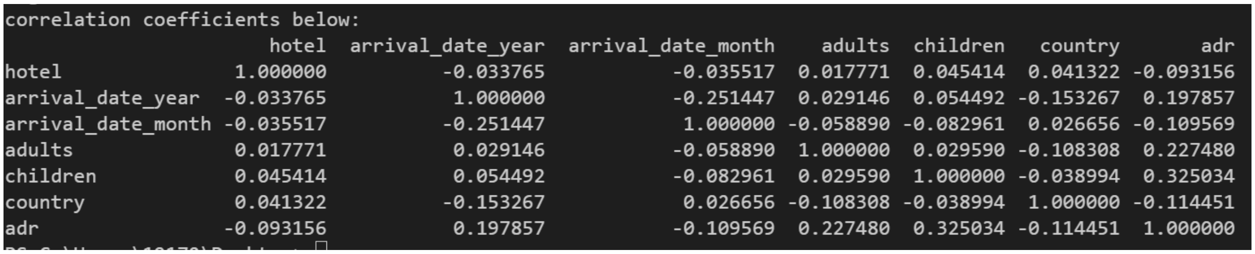


```
--------------------------
[0]
Year of Arrival:  2015
Month of Arrival:  2
Number of Adults:  1
Number of Children:  0
Country:  48
Average Daily Rate:  51
```

As a result the next person to book a hotel will book a city hotel on December 2015 at Spain where there will be one adult and zero children and the average daily rate is $51.

OLS Regression Model 1 (with three variables)

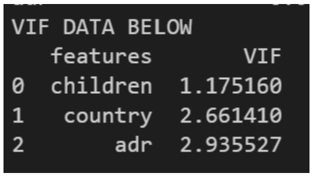
When working with Regression models, we must first find which variables affect the target variable y, hotel type chosen, the most. To do so, we must analyze the correlation coefficients below:



```
correlation coefficients below:
                      hotel  arrival_date_year  arrival_date_month    adults  children   country       adr
hotel              1.000000          -0.033765           -0.035517  0.017771  0.045414  0.041322 -0.093156
arrival_date_year  -0.033765          1.000000           -0.251447  0.029146  0.054492 -0.153267  0.197857
arrival_date_month -0.035517         -0.251447            1.000000 -0.058890 -0.082961  0.026656 -0.109569
adults             0.017771           0.029146           -0.058890  1.000000  0.029590 -0.108308  0.227480
children           0.045414           0.054492           -0.082961  0.029590  1.000000 -0.038994  0.325034
country            0.041322          -0.153267            0.026656 -0.108308 -0.038994  1.000000 -0.114451
adr               -0.093156           0.197857           -0.109569  0.227480  0.325034 -0.114451  1.000000
```

From this we can see that the top three correlations with hotel are children, country, and adr.

Now we will check the variance inflation factor (VIF) for these columns

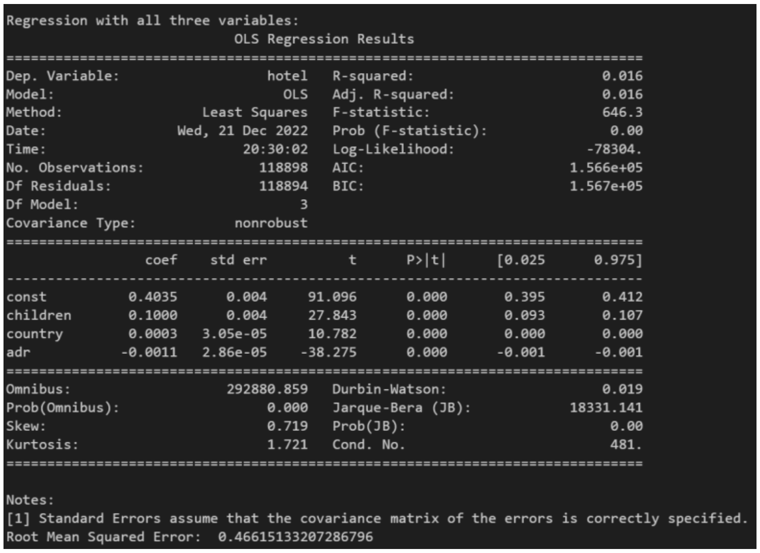


```
VIF DATA BELOW
   features       VIF
0  children  1.175160
1   country  2.661410
2       adr  2.935527
```

The VIF measures the amount of variance of the coefficient derived from the model is inflated by collinearity. Since the values for VIF are not 5-10, we don't need to drop any of the variables, and we shouldn't have any problems with collinearity.

Now we will check the ordinary least square models (OLS) for these columns

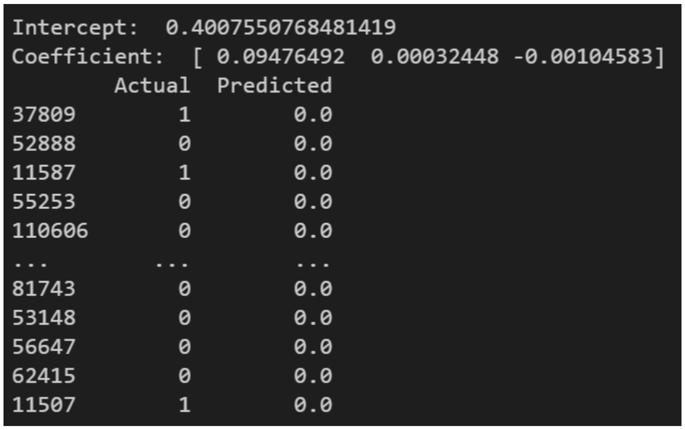
<u>MODEL 1 with three variables: children, country, and adr</u>



```
Regression with all three variables:
                            OLS Regression Results
==============================================================================
Dep. Variable:                  hotel   R-squared:                       0.016
Model:                            OLS   Adj. R-squared:                  0.016
Method:                 Least Squares   F-statistic:                     646.3
Date:                Wed, 21 Dec 2022   Prob (F-statistic):               0.00
Time:                        20:30:02   Log-Likelihood:                -78304.
No. Observations:              118898   AIC:                         1.566e+05
Df Residuals:                  118894   BIC:                         1.567e+05
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.4035      0.004     91.096      0.000       0.395       0.412
children       0.1000      0.004     27.843      0.000       0.093       0.107
country        0.0003   3.05e-05     10.782      0.000       0.000       0.000
adr           -0.0011   2.86e-05    -38.275      0.000      -0.001      -0.001
==============================================================================
Omnibus:                   292880.859   Durbin-Watson:                   0.019
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18331.141
Skew:                           0.719   Prob(JB):                         0.00
Kurtosis:                       1.721   Cond. No.                         481.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Root Mean Squared Error:  0.46615133207286796
```

From these results, we see that the value for R-squared(1.6%) and adjusted r squared(1.6%) is very low. The root mean square error also seems to be low for this model, which is a good sign.

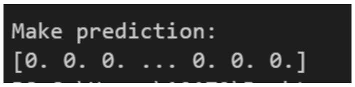
When we continue to use this model we achieve the results:



```
Intercept:  0.4007550768481419
Coefficient:  [ 0.09476492  0.00032448 -0.00104583]
        Actual  Predicted
37809        1        0.0
52888        0        0.0
11587        1        0.0
55253        0        0.0
110606       0        0.0
...        ...        ...
81743        0        0.0
53148        0        0.0
56647        0        0.0
62415        0        0.0
11507        1        0.0
```

From the image above, we can compare how well the model predicts whether a person chooses a resort hotel or city hotel compared to the real dataset. You may notice that most of the predicted values match the actual values which is a good sign that this model is good.

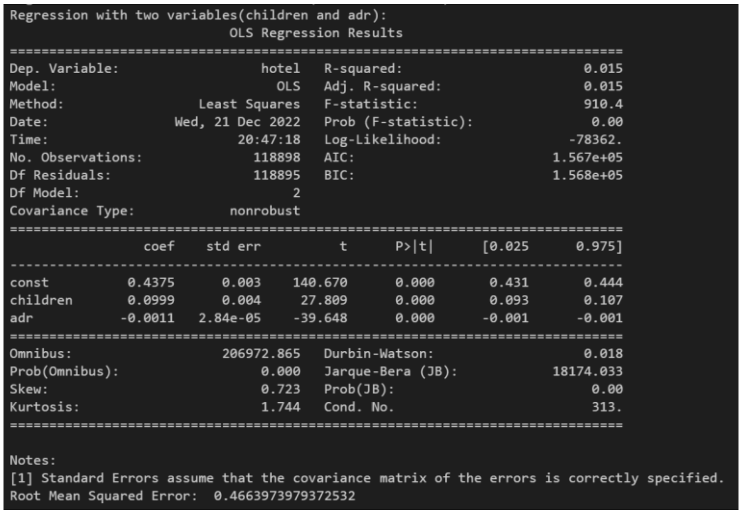
As a result, the prediction of the linear regression model with the use of three variables, country, children, and adr predicted that the next set of booking will consist mainly of bookings for city hotels as follows:



```
Make prediction:
[0. 0. 0. ... 0. 0. 0.]
```

OLS Regression Model 2 (with two variables)

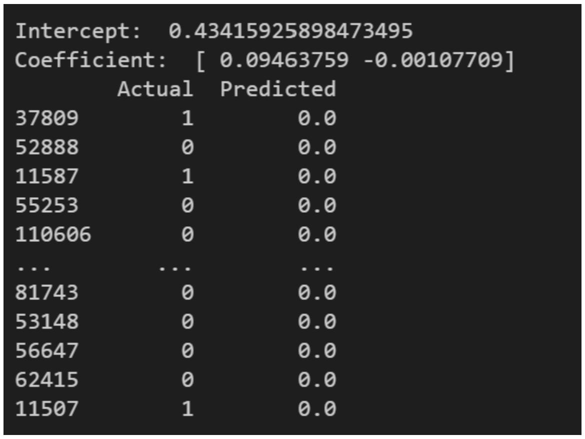
Now let’s try dropping the country column for the regression model since it was the least amongst the three correlation coefficient values. Now we will try creating a model with two variables children and adr.



```
Regression with two variables(children and adr):
                            OLS Regression Results
==============================================================================
Dep. Variable:                  hotel   R-squared:                       0.015
Model:                            OLS   Adj. R-squared:                  0.015
Method:                 Least Squares   F-statistic:                     910.4
Date:                Wed, 21 Dec 2022   Prob (F-statistic):               0.00
Time:                        20:47:18   Log-Likelihood:                -78362.
No. Observations:              118898   AIC:                         1.567e+05
Df Residuals:                  118895   BIC:                         1.568e+05
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.4375      0.003    140.670      0.000       0.431       0.444
children       0.0999      0.004     27.809      0.000       0.093       0.107
adr           -0.0011   2.84e-05    -39.648      0.000      -0.001      -0.001
==============================================================================
Omnibus:                   206972.865   Durbin-Watson:                   0.018
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18174.033
Skew:                           0.723   Prob(JB):                         0.00
Kurtosis:                       1.744   Cond. No.                         313.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Root Mean Squared Error:  0.4663973979372532
```

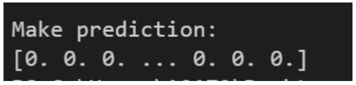
Compared to the past model the r-squared value and adjusted r-squared value of this model is 0.1% lower. Additionally, the root mean square error is around the same but has gone up in this model indicating model 1 is very slightly better than model 2.

When we continue to use this model we achieve the results:



```
Intercept:  0.43415925898473495
Coefficient:  [ 0.09463759 -0.00107709]
        Actual  Predicted
37809        1        0.0
52888        0        0.0
11587        1        0.0
55253        0        0.0
110606       0        0.0
...        ...        ...
81743        0        0.0
53148        0        0.0
56647        0        0.0
62415        0        0.0
11507        1        0.0
```

From the image above, we can see how well the model predicts whether a person chooses a resort hotel or city hotel compared to the real dataset. You may notice that just like the regression model with three variables, most of the predicted values match the actual values which is a good sign that this model is good. As a result, the prediction of the linear regression model with the use of two variables children and adr predicted that the next set of booking will consist mainly of bookings for city hotels as follows:



```
Make prediction:
[0. 0. 0. ... 0. 0. 0.]
```

Conclusion

The three models created for this study were the Gaussian model, Linear Regression model with three variables, and Linear regression with two variables. Although both Linear Regression models were similar, the best model produced seemed to be the Linear Regression Model with the three variables (children, country, and adr). After exploring the data, it was visible that city hotels were mainly preferred by most people booking hotels in the dataset.

All three models produced the same prediction of the next bookings to be for city hotels. These predictions seemed to be very accurate after observing the data and seeing that most of the bookings except for the ones with kids in 2015 favored city hotels over resort hotels. When using the Gaussian Model to make a prediction, the prediction predicted a city hotel booking in December. From this dataset, we can conclude that city hotels have become the more popular choice for hotel over resort hotel even during the summer months.

From this study, we can conclude that most bookings whether it be with children or only adults choose city hotels and book during August and July. As a result, resorts should find ways they can attract their customers again. Since resorts were popular amongst bookings with children during 2015, perhaps resorts can find ways to cater towards children. Additionally, this study suggests that people may find it harder to book hotels during August and July due to the high demand amongst both bookings for only adults and bookings with children.

Resources

Mostipak, J. (2020, February 13). *Hotel Booking Demand*. Kaggle. Retrieved from https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand?resource=download