Michael Kojo Ampah

Machine-Learning

Booking of hotel

In [22]:
```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import random as rd
from sklearn import datasets
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [25]:
```
data= pd.read_csv(r'C:\Users\M_Ampah\Downloads\booking_of_hotel.csv')
```

Exploratory Data Analysis

In [26]:
```
data.head()
```

Out[26]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_
0	Resort Hotel	0	342	2015	July	
1	Resort Hotel	0	737	2015	July	
2	Resort Hotel	0	7	2015	July	
3	Resort Hotel	0	13	2015	July	
4	Resort Hotel	0	14	2015	July	

5 rows × 32 columns

In [27]:
```
data.tail()
```

Out[27]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_w
119385	City Hotel	0	23	2017	August	
119386	City Hotel	0	102	2017	August	
119387	City Hotel	0	34	2017	August	
119388	City Hotel	0	109	2017	August	
119389	City Hotel	0	205	2017	August	

5 rows × 32 columns

In [29]:
```
#No of Rows and columns
data.shape
```

Out[29]:
```
(119390, 32)


119390 rows and 32 columns
```

In [30]:
```
# All columns in the dataset
data.columns
```

Out[30]:
```
Index(['hotel', 'is_canceled', 'lead_time', 'arrival_date_year',
       'arrival_date_month', 'arrival_date_week_number',
       'arrival_date_day_of_month', 'stays_in_weekend_nights',
       'stays_in_week_nights', 'adults', 'children', 'babies', 'meal',
       'country', 'market_segment', 'distribution_channel',
       'is_repeated_guest', 'previous_cancellations',
       'previous_bookings_not_canceled', 'reserved_room_type',
       'assigned_room_type', 'booking_changes', 'deposit_type', 'agen
t',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [61]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119390 entries, 0 to 119389
Data columns (total 32 columns):
 #   Column                          Non-Null Count   Dtype
---  ------                          --------------   -----
 0   hotel                           119390 non-null  object
 1   is_canceled                     119390 non-null  int64
 2   lead_time                       119390 non-null  int64
 3   arrival_date_year               119390 non-null  int64
 4   arrival_date_month              119390 non-null  object
 5   arrival_date_week_number        119390 non-null  int64
 6   arrival_date_day_of_month       119390 non-null  int64
 7   stays_in_weekend_nights         119390 non-null  int64
 8   stays_in_week_nights            119390 non-null  int64
 9   adults                          119390 non-null  int64
 10  children                        119386 non-null  float64
 11  babies                          119390 non-null  int64
 12  meal                            119390 non-null  object
 13  country                         118902 non-null  object
 14  market_segment                  119390 non-null  object
 15  distribution_channel            119390 non-null  object
 16  is_repeated_guest               119390 non-null  int64
 17  previous_cancellations          119390 non-null  int64
 18  previous_bookings_not_canceled  119390 non-null  int64
 19  reserved_room_type              119390 non-null  object
 20  assigned_room_type              119390 non-null  object
 21  booking_changes                 119390 non-null  int64
 22  deposit_type                    119390 non-null  object
 23  agent                           103050 non-null  float64
 24  company                         6797 non-null    float64
 25  days_in_waiting_list            119390 non-null  int64
 26  customer_type                   119390 non-null  object
 27  adr                             119390 non-null  float64
 28  required_car_parking_spaces     119390 non-null  int64
 29  total_of_special_requests       119390 non-null  int64
 30  reservation_status              119390 non-null  object
 31  reservation_status_date         119390 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(4), int64(16), object(11)
memory usage: 29.1+ MB
```

In [66]:
```
data.isnull().sum() # Checking the null value
```

Out[66]:
```
hotel                           0
is_canceled                     0
lead_time                       0
arrival_date_year               0
arrival_date_month              0
arrival_date_week_number        0
arrival_date_day_of_month       0
stays_in_weekend_nights         0
stays_in_week_nights            0
adults                          0
children                        0
babies                          0
meal                            0
country                         0
market_segment                  0
distribution_channel            0
is_repeated_guest               0
previous_cancellations          0
previous_bookings_not_canceled  0
reserved_room_type              0
assigned_room_type              0
booking_changes                 0
deposit_type                    0
days_in_waiting_list            0
customer_type                   0
adr                             0
required_car_parking_spaces     0
total_of_special_requests       0
reservation_status              0
reservation_status_date         0
dtype: int64
```

In [59]:
```
data.describe(include = 'object')
```

Out[59]:

	hotel	arrival_date_month	meal	country	market_segment	distribution_channel
count	119390	119390	119390	118902	119390	119390
unique	2	12	5	177	8	5
top	City Hotel	August	BB	PRT	Online TA	TA/TO
freq	79330	13877	92310	48590	56477	97870

In [41]:
```
data.describe()
```

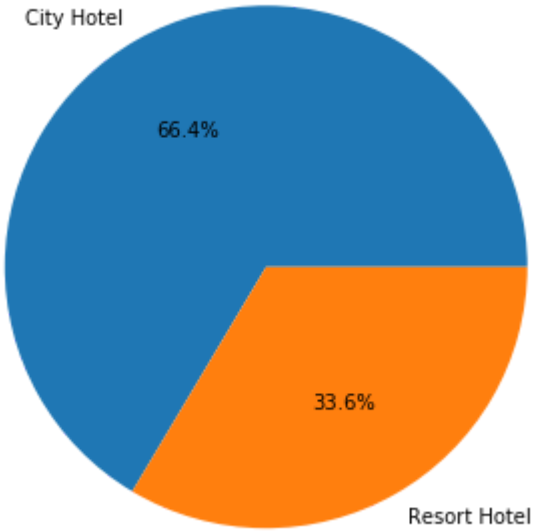
Out[41]:

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_
count	119390.000000	119390.000000	119390.000000	119390.000000	
mean	0.370416	104.011416	2016.156554	27.165173	
std	0.482918	106.863097	0.707476	13.605138	
min	0.000000	0.000000	2015.000000	1.000000	
25%	0.000000	18.000000	2016.000000	16.000000	
50%	0.000000	69.000000	2016.000000	28.000000	
75%	1.000000	160.000000	2017.000000	38.000000	
max	1.000000	737.000000	2017.000000	53.000000	

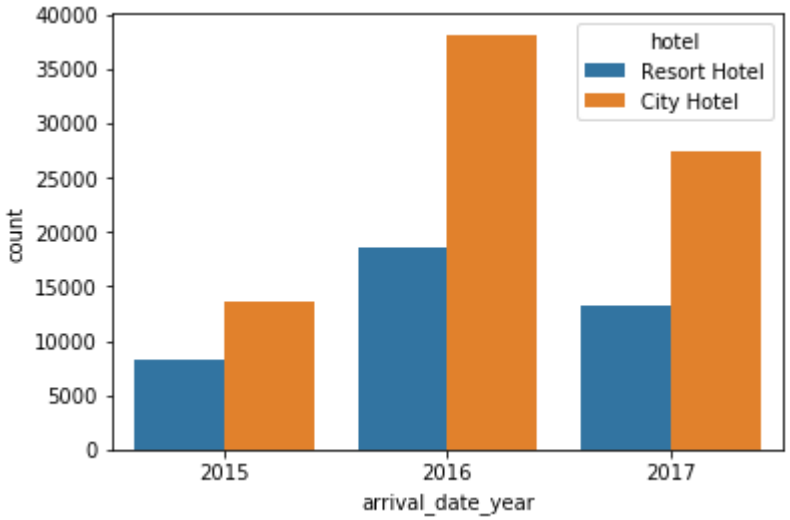
Data Analysis and Visualizations

In [50]:
```
#Reservation status Graph (canceled and not canceled)
plt.figure(figsize = (4,4))
plt.title('Reservation status')
plt.bar(['Not canceled','canceled'],data['is_canceled'].value_counts(),
plt.show()
```

In [77]:
```
# Plotting the value counts of hotel column
l=['City Hotel','Resort Hotel']
plt.figure(figsize = (8,6))
plt.pie(data["hotel"].value_counts(),labels=l,autopct="%0.01f%%")
plt.show()
```



In [81]:
```
#Arrival Years
sns.countplot(data["arrival_date_year"],hue=data["hotel"]);
```



In [ ]: