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

Read the data

In [2]:
```
df = pd.read_csv(r"D:\datasets\Day3/hotel_bookings.csv")
```

In [3]:
```
df.head()
```

Out[3]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_
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 [4]:
```
df.shape
```

Out[4]:
```
(119390, 32)
```

In [5]:
```
df.dtypes
```

Out[5]:
```
hotel                            object
is_canceled                       int64
lead_time                         int64
arrival_date_year                 int64
arrival_date_month               object
arrival_date_week_number          int64
arrival_date_day_of_month         int64
stays_in_weekend_nights           int64
stays_in_week_nights              int64
adults                            int64
children                        float64
babies                            int64
meal                             object
country                          object
market_segment                   object
distribution_channel             object
is_repeated_guest                 int64
previous_cancellations            int64
previous_bookings_not_canceled    int64
reserved_room_type               object
assigned_room_type               object
booking_changes                   int64
deposit_type                     object
agent                           float64
company                         float64
days_in_waiting_list              int64
customer_type                    object
adr                             float64
required_car_parking_spaces       int64
total_of_special_requests         int64
reservation_status               object
reservation_status_date          object
dtype: object
```

Doing Data Cleaning

In [6]:
```
df.columns
```

Out[6]:
```
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', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [7]:
```
filter1 = (df['children'] == 0) & (df['adults'] == 0) & (df['babies'] == 0)
```

In [8]:
```
df[filter1]
```

Out[8]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	ar
2224	Resort Hotel	0	1	2015	October	
2409	Resort Hotel	0	0	2015	October	
3181	Resort Hotel	0	36	2015	November	
3684	Resort Hotel	0	165	2015	December	
3708	Resort Hotel	0	165	2015	December	
...	...	...	...	...	...	
115029	City Hotel	0	107	2017	June	
115091	City Hotel	0	1	2017	June	
116251	City Hotel	0	44	2017	July	
116534	City Hotel	0	2	2017	July	
117087	City Hotel	0	170	2017	July	

180 rows × 32 columns

In [9]:
```
df.shape
```

Out[9]:
```
(119390, 32)
```

In [10]:
```
df[filter1].shape
```

Out[10]:
```
(180, 32)
```

In [11]:
```
df[~filter1]
```

Out[11]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	ar
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	
...	...	...	...	...	...	
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	

119210 rows × 32 columns

In [12]:
```
df2= df[~filter1]
```

In [13]:
```
df2.duplicated()
```

Out[13]:
```
0         False
1         False
2         False
3         False
4         False
          ...
119385    False
119386    False
119387    False
119388    False
119389    False
Length: 119210, dtype: bool
```

In [14]:
```
df2.duplicated().sum()
```

Out[14]:
```
np.int64(31980)
```

In [15]:
```
df2.shape
```

Out[15]:
```
(119210, 32)
```

In [16]:
```
data = df2.drop_duplicates()
```

In [17]:
```
data.shape
```

Out[17]:
```
(87230, 32)
```

Performing descriptive analysis

In [18]:
```
data.columns
```

Out[18]:
```
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', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [19]:
```
data[['lead_time' , 'total_of_special_requests','adr']].describe().T
```

Out[19]:

	count	mean	std	min	25%	50%	75%
lead_time	87230.0	79.971019	86.058683	0.00	11.00	49.0	125.
total_of_special_requests	87230.0	0.698934	0.832051	0.00	0.00	0.0	1.
adr	87230.0	106.518031	54.891227	-6.38	72.25	98.2	134.

In [20]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 87230 entries, 0 to 119389
Data columns (total 32 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   hotel                           87230 non-null  object 
 1   is_canceled                     87230 non-null  int64  
 2   lead_time                       87230 non-null  int64  
 3   arrival_date_year               87230 non-null  int64  
 4   arrival_date_month              87230 non-null  object 
 5   arrival_date_week_number        87230 non-null  int64  
 6   arrival_date_day_of_month       87230 non-null  int64  
 7   stays_in_weekend_nights         87230 non-null  int64  
 8   stays_in_week_nights            87230 non-null  int64  
 9   adults                          87230 non-null  int64  
 10  children                        87226 non-null  float64
 11  babies                          87230 non-null  int64  
 12  meal                            87230 non-null  object 
 13  country                         86783 non-null  object 
 14  market_segment                  87230 non-null  object 
 15  distribution_channel            87230 non-null  object 
 16  is_repeated_guest               87230 non-null  int64  
 17  previous_cancellations          87230 non-null  int64  
 18  previous_bookings_not_canceled  87230 non-null  int64  
 19  reserved_room_type              87230 non-null  object 
 20  assigned_room_type              87230 non-null  object 
 21  booking_changes                 87230 non-null  int64  
 22  deposit_type                    87230 non-null  object 
 23  agent                           75089 non-null  float64
 24  company                         5237 non-null   float64
 25  days_in_waiting_list            87230 non-null  int64  
 26  customer_type                   87230 non-null  object 
 27  adr                             87230 non-null  float64
 28  required_car_parking_spaces     87230 non-null  int64  
 29  total_of_special_requests       87230 non-null  int64  
 30  reservation_status              87230 non-null  object 
 31  reservation_status_date         87230 non-null  object 
dtypes: float64(4), int64(16), object(12)
memory usage: 22.0+ MB
```

In [21]:
```
data.info(memory_usage = 'deep')
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 87230 entries, 0 to 119389
Data columns (total 32 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   hotel                           87230 non-null  object 
 1   is_canceled                     87230 non-null  int64  
 2   lead_time                       87230 non-null  int64  
 3   arrival_date_year               87230 non-null  int64  
 4   arrival_date_month              87230 non-null  object 
 5   arrival_date_week_number        87230 non-null  int64  
 6   arrival_date_day_of_month       87230 non-null  int64  
 7   stays_in_weekend_nights         87230 non-null  int64  
 8   stays_in_week_nights            87230 non-null  int64  
 9   adults                          87230 non-null  int64  
 10  children                        87226 non-null  float64
 11  babies                          87230 non-null  int64  
 12  meal                            87230 non-null  object 
 13  country                         86783 non-null  object 
 14  market_segment                  87230 non-null  object 
 15  distribution_channel            87230 non-null  object 
 16  is_repeated_guest               87230 non-null  int64  
 17  previous_cancellations          87230 non-null  int64  
 18  previous_bookings_not_canceled  87230 non-null  int64  
 19  reserved_room_type              87230 non-null  object 
 20  assigned_room_type              87230 non-null  object 
 21  booking_changes                 87230 non-null  int64  
 22  deposit_type                    87230 non-null  object 
 23  agent                           75089 non-null  float64
 24  company                         5237 non-null   float64
 25  days_in_waiting_list            87230 non-null  int64  
 26  customer_type                   87230 non-null  object 
 27  adr                             87230 non-null  float64
 28  required_car_parking_spaces     87230 non-null  int64  
 29  total_of_special_requests       87230 non-null  int64  
 30  reservation_status              87230 non-null  object 
 31  reservation_status_date         87230 non-null  object 
dtypes: float64(4), int64(16), object(12)
memory usage: 69.2 MB
```

In [22]:
```
for col in ['lead_time','total_of_special_requests','adr']:
    print('feature name : {}'.format(col))
```

```
feature name : lead_time
feature name : total_of_special_requests
feature name : adr
```

In [23]:
```
for col in ['lead_time','total_of_special_requests','adr']:
    print('feature name: {}'.format(col))

    for i in range(90,101,1):
        quantile_value = np.quantile(data[col] , q= 1/100)
        print('{}th quantile value is {}'.format(i, quantile_value))
    print('\n')
```

```
feature name: lead_time
90th quantile value is 0.0
91th quantile value is 0.0
92th quantile value is 0.0
93th quantile value is 0.0
94th quantile value is 0.0
95th quantile value is 0.0
96th quantile value is 0.0
97th quantile value is 0.0
98th quantile value is 0.0
99th quantile value is 0.0
100th quantile value is 0.0


feature name: total_of_special_requests
90th quantile value is 0.0
91th quantile value is 0.0
92th quantile value is 0.0
93th quantile value is 0.0
94th quantile value is 0.0
95th quantile value is 0.0
96th quantile value is 0.0
97th quantile value is 0.0
98th quantile value is 0.0
99th quantile value is 0.0
100th quantile value is 0.0


feature name: adr
90th quantile value is 0.0
91th quantile value is 0.0
92th quantile value is 0.0
93th quantile value is 0.0
94th quantile value is 0.0
95th quantile value is 0.0
96th quantile value is 0.0
97th quantile value is 0.0
98th quantile value is 0.0
99th quantile value is 0.0
100th quantile value is 0.0
```

Spatial Analysis on Guests Home-Town

In [24]:
```
df.head()
```

Out[24]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_
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 [25]:
```
df.columns
```

Out[25]:
```
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', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [26]:
```
not_cancelled = data[data['is_canceled'] == 0 ]
```

In [27]:
```
not_cancelled.head(3)
```

Out[27]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_
0	Resort Hotel	0	342	2015	July	
1	Resort Hotel	0	737	2015	July	
2	Resort Hotel	0	7	2015	July	

3 rows × 32 columns

In [28]:
```
country_wise_data = not_cancelled['country'].value_counts().reset_index()
```

In [29]:
```
country_wise_data
```

Out[29]:

	country	count
0	PRT	17573
1	GBR	8440
2	FRA	7091
3	ESP	5382
4	DEU	4332
...	...	...
160	KIR	1
161	ATF	1
162	TJK	1
163	SLE	1
164	FRO	1

165 rows × 2 columns

In [30]:
```
country_wise_data.columns = ['country','No of guests']
```

In [31]:
```
country_wise_data
```

Out[31]:

	country	No of guests
0	PRT	17573
1	GBR	8440
2	FRA	7091
3	ESP	5382
4	DEU	4332
...	...	...
160	KIR	1
161	ATF	1
162	TJK	1
163	SLE	1
164	FRO	1

165 rows × 2 columns

In [32]:
```
!pip install chart-studio
!pip install plotly
```

```
Requirement already satisfied: chart-studio in c:\programdata\anaconda3\lib\sit
e-packages (1.1.0)
Requirement already satisfied: plotly in c:\programdata\anaconda3\lib\site-pack
ages (from chart-studio) (5.24.1)
Requirement already satisfied: requests in c:\programdata\anaconda3\lib\site-pa
ckages (from chart-studio) (2.32.3)
Requirement already satisfied: retrying>=1.3.3 in c:\programdata\anaconda3\lib\
site-packages (from chart-studio) (1.4.1)
Requirement already satisfied: six in c:\users\administrator\appdata\roaming\py
thon\python313\site-packages (from chart-studio) (1.17.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\programdata\anaconda3\lib\
site-packages (from plotly->chart-studio) (9.0.0)
Requirement already satisfied: packaging in c:\users\administrator\appdata\roam
ing\python\python313\site-packages (from plotly->chart-studio) (25.0)
Requirement already satisfied: charset-normalizer<4,>=2 in c:\programdata\anaco
nda3\lib\site-packages (from requests->chart-studio) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in c:\programdata\anaconda3\lib\sit
e-packages (from requests->chart-studio) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in c:\programdata\anaconda3\l
ib\site-packages (from requests->chart-studio) (2.3.0)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\l
ib\site-packages (from requests->chart-studio) (2025.4.26)
Requirement already satisfied: plotly in c:\programdata\anaconda3\lib\site-pack
ages (5.24.1)
Requirement already satisfied: tenacity>=6.2.0 in c:\programdata\anaconda3\lib\
site-packages (from plotly) (9.0.0)
Requirement already satisfied: packaging in c:\users\administrator\appdata\roam
ing\python\python313\site-packages (from plotly) (25.0)
```

In [33]:
```
import chart_studio.plotly as py
import plotly.graph_objs as go
import plotly.express as px
from plotly.offline import download_plotlyjs , init_notebook_mode , plot , ipl

init_notebook_mode(connected = True)
```

In [34]:
```
map_guest = px.choropleth(data_frame = country_wise_data,
                          locations = country_wise_data['country'],
                          color = country_wise_data['No of guests'],
                          hover_name=country_wise_data['country'],
                          title = "Home country of Guests",
                          width = 1000,
                          height = 700
                         )
```

In [35]:
```
map_guest.show()
```

Analysing difference between assigned and reserved room types

In [36]:
```
data.columns
```

Out[36]:
```
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', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [37]:
```
pivot = pd.crosstab(index = data['reserved_room_type'] , columns = data['assig
```

In [38]:
```
pivot
```

Out[38]:

assigned_room_type	A	B	C	D	E	F	G	H	I	K
reserved_room_type										
A	45850	892	1253	6402	1034	390	176	94	205	140
B	106	872	0	5	2	2	8	0	0	1
C	5	2	866	6	4	2	10	9	10	0
D	295	27	32	15979	657	199	82	9	67	29
E	15	2	6	22	5458	383	97	4	40	9
F	6	14	0	4	31	2636	113	3	10	3
G	5	1	2	0	4	14	1999	7	15	3
H	0	0	0	1	0	0	10	579	6	0
L	1	1	1	0	0	1	0	1	0	0
All	46283	1811	2160	22419	7190	3627	2495	706	353	185

In [39]:
```
pivot_normalize = pd.crosstab(index = data['reserved_room_type'] , columns = d
```

In [41]:
```
pivot_normalize
```

Out[41]:

assigned_room_type reserved_room_type	A	B	C	D	E	F	G	H	I	K	L
A	81.0	2.0	2.0	11.0	2.0	1.0	0.0	0.0	0.0	0.0	0.0
B	11.0	88.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
C	1.0	0.0	95.0	1.0	0.0	0.0	1.0	1.0	1.0	0.0	0.0
D	2.0	0.0	0.0	92.0	4.0	1.0	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	0.0	90.0	6.0	2.0	0.0	1.0	0.0	0.0
F	0.0	0.0	0.0	0.0	1.0	93.0	4.0	0.0	0.0	0.0	0.0
G	0.0	0.0	0.0	0.0	0.0	1.0	98.0	0.0	1.0	0.0	0.0
H	0.0	0.0	0.0	0.0	0.0	0.0	2.0	97.0	1.0	0.0	0.0
L	17.0	17.0	17.0	0.0	0.0	17.0	0.0	17.0	0.0	0.0	17.0
All	53.0	2.0	2.0	26.0	8.0	4.0	3.0	1.0	0.0	0.0	0.0

Which market segment has highest bookings?

In [42]:
```
data['market_segment'].value_counts()
```

Out[42]:
```
market_segment
Online TA        51553
Offline TA/TO    13855
Direct           11780
Groups            4922
Corporate         4200
Complementary      692
Aviation           226
Undefined            2
Name: count, dtype: int64
```

In [43]:
```
data['market_segment'].value_counts().values
```

Out[43]:
```
array([51553, 13855, 11780,  4922,  4200,   692,   226,     2])
```

In [44]:
```
data['market_segment'].value_counts().index
```

Out[44]:
```
Index(['Online TA', 'Offline TA/TO', 'Direct', 'Groups', 'Corporate',
       'Complementary', 'Aviation', 'Undefined'],
      dtype='object', name='market_segment')
```

In [46]:
```
fig = px.pie(data,
       values = data['market_segment'].value_counts().values,
       names = data['market_segment'].value_counts().index)
```

In [47]:
```
fig.show()
```

Analysing average price per night of various room-types for all the market segment

In [48]:
```
data.columns
```

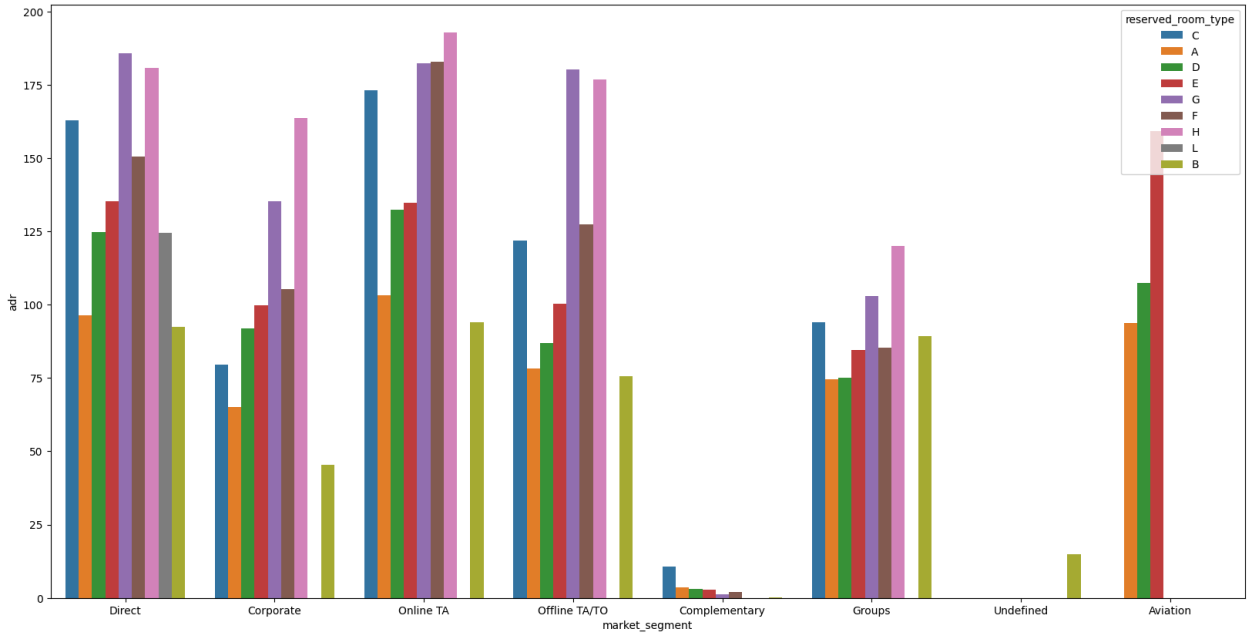
Out[48]:
```
Index(['hotel', 'is_canceled', 'lead_time', 'arrival_date_year',
       'arrival_date_month', 'arrival_date_week_number',
       'arrival_date_day_of_month', 'stays_in_weekend_nights',
       'stays_in_week_nights', 'adults', 'children', 'babies', 'meal',
       'country', 'market_segment', 'distribution_channel',
       'is_repeated_guest', 'previous_cancellations',
       'previous_bookings_not_canceled', 'reserved_room_type',
       'assigned_room_type', 'booking_changes', 'deposit_type', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date'],
      dtype='object')
```

In [52]:
```
plt.figure(figsize = (20, 10))
sns.barplot(x="market_segment" , y= "adr" , hue = "reserved_room_type" , data
```

```
C:\Users\Administrator\AppData\Local\Temp\ipykernel_24676\3408636515.py:2: Futu
reWarning:


The `ci` parameter is deprecated. Use `errorbar=None` for the same effect.
```

Out[52]: <Axes: xlabel='market_segment', ylabel='adr'>



Analysing distribution of guest arrival

In [53]:
```
data.head()
```

Out[53]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_
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 [55]:
```
data['arrival_date_month'].unique()
```

Out[55]:
```
array(['July', 'August', 'September', 'October', 'November', 'December',
       'January', 'February', 'March', 'April', 'May', 'June'],
      dtype=object)
```

In [56]:
```
dict_month = {'July':7, 'August':8, 'September':9, 'October':10, 'November':11
       'January':1, 'February':2, 'March':3, 'April':4, 'May':5, 'June':6}
```

In [58]:
```
import warnings
from warnings import filterwarnings
filterwarnings('ignore')
```

In [59]:
```
data['arrival_date_month_index'] = data['arrival_date_month'].map(dict_month)
```

In [60]:
```
data.columns
```

Out[60]:
```
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', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date',
       'arrival_date_month_index'],
      dtype='object')
```

In [61]:
```
data[['arrival_date_year','arrival_date_month_index','arrival_date_day_of_mont
```

Out[61]:

	arrival_date_year	arrival_date_month_index	arrival_date_day_of_month
0	2015	7	1
1	2015	7	1
2	2015	7	1
3	2015	7	1
4	2015	7	1
...	...	...	...
119385	2017	8	30
119386	2017	8	31
119387	2017	8	31
119388	2017	8	31
119389	2017	8	29

87230 rows × 3 columns

In [64]:
```
data['arrival_date']= data['arrival_date_year'].astype(str)+'-'+data['arrival_
```

In [65]:
```
data.head()
```

Out[65]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_
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 × 34 columns

In [66]:
```
data.columns
```

Out[66]:
```
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', 'agent',
       'company', 'days_in_waiting_list', 'customer_type', 'adr',
       'required_car_parking_spaces', 'total_of_special_requests',
       'reservation_status', 'reservation_status_date',
       'arrival_date_month_index', 'arrival_date'],
      dtype='object')
```

In [67]:
```
data[['adults', 'children', 'babies']]
```

Out[67]:

	adults	children	babies
0	2	0.0	0
1	2	0.0	0
2	1	0.0	0
3	1	0.0	0
4	2	0.0	0
...	...	...	...
119385	2	0.0	0
119386	3	0.0	0
119387	2	0.0	0
119388	2	0.0	0
119389	2	0.0	0

87230 rows × 3 columns

In [68]:
```
data['Total_guests'] = data['adults'] + data['children'] + data['babies']
```

In [69]:
```
data.head(3)
```

Out[69]:

	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_
0	Resort Hotel	0	342	2015	July	
1	Resort Hotel	0	737	2015	July	
2	Resort Hotel	0	7	2015	July	

3 rows × 35 columns

In [70]:
```
data[['arrival_date','Total_guests']]
```

Out[70]:

	arrival_date	Total_guests
0	2015-7-1	2.0
1	2015-7-1	2.0
2	2015-7-1	1.0
3	2015-7-1	1.0
4	2015-7-1	2.0
...	...	...
119385	2017-8-30	2.0
119386	2017-8-31	3.0
119387	2017-8-31	2.0
119388	2017-8-31	2.0
119389	2017-8-29	2.0

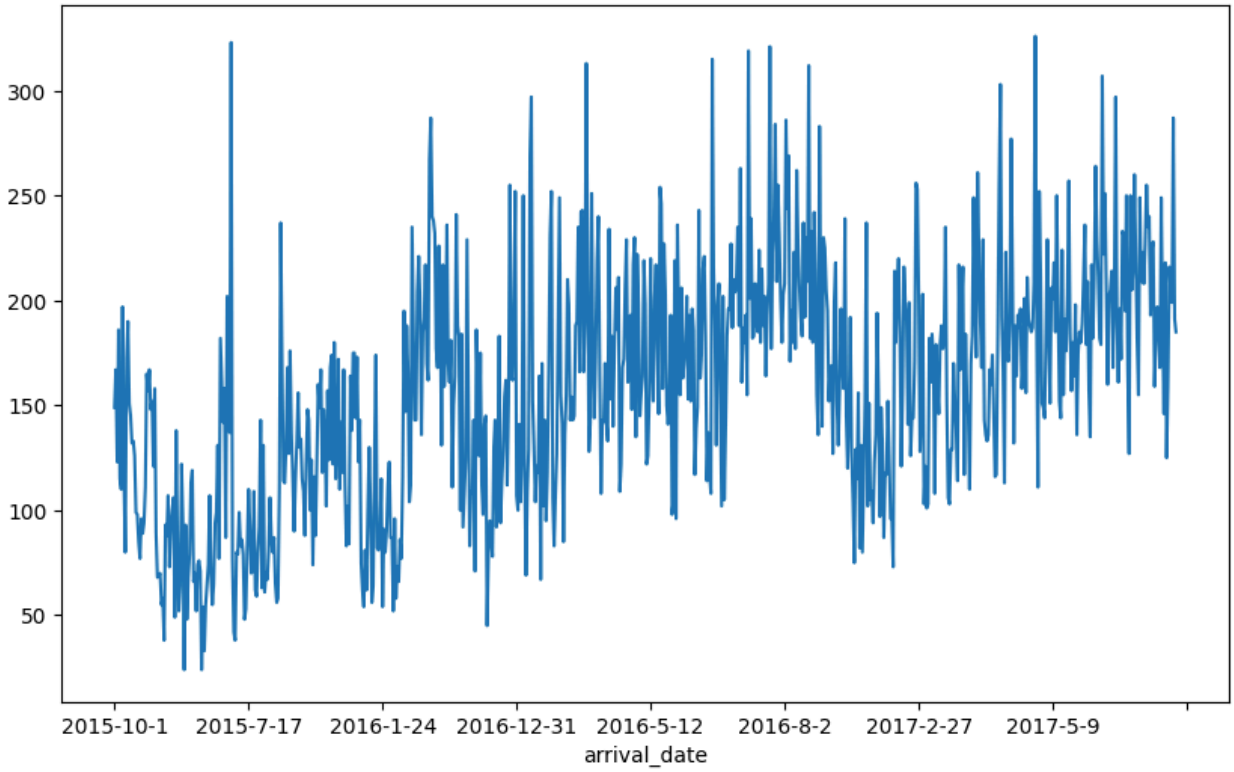
87230 rows × 2 columns

In [71]:
```
dataNoCancel = data[data['is_canceled'] == 0 ]
```

In [76]:
```
guest_arrival_series = dataNoCancel.groupby(['arrival_date'])['Total_guests'].
```

In [79]:
```
guest_arrival_series.plot(figsize = (10,6))
```

Out[79]:
```
<Axes: xlabel='arrival_date'>
```



In [ ]:

In [80]:
```
guest_arrival_series
```

Out[80]:
```
arrival_date
2015-10-1     149.0
2015-10-10    167.0
2015-10-11    123.0
2015-10-12    186.0
2015-10-13    115.0
              ...
2017-8-5      205.0
2017-8-6      199.0
2017-8-7      287.0
2017-8-8      191.0
2017-8-9      185.0
Name: Total_guests, Length: 793, dtype: float64
```

In [81]:
```
type(guest_arrival_series)
```

Out[81]:
```
pandas.core.series.Series
```

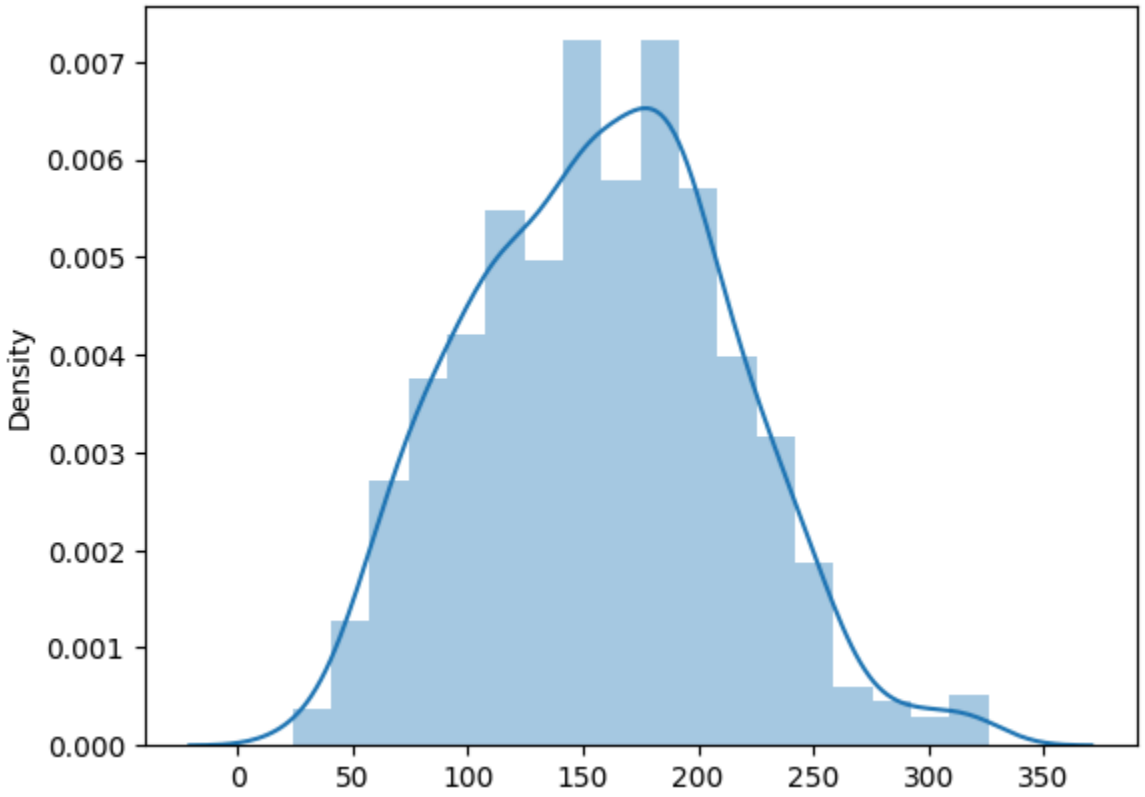
In [82]:
```
guest_arrival_series.values
```

Out[82]:
```
array([149., 167., 123., 186., 115., 110., 197., 118.,  80., 172., 190.,
       151., 145., 132., 133., 126.,  99.,  98.,  85.,  77.,  96.,  89.,
        94., 111., 165., 157., 167., 148., 152., 121., 158.,  89.,  68.,
        69.,  70.,  55.,  58.,  38.,  93.,  88., 107.,  73.,  92., 100.,
       106.,  49., 138.,  81.,  52.,  64., 122.,  95.,  24.,  93.,  48.,
        68.,  79., 113., 119.,  66.,  70.,  52.,  74.,  76.,  71.,  24.,
        54.,  33.,  55.,  64.,  74., 107.,  68.,  55.,  65.,  94.,  99.,
       131.,  77., 182., 161., 142., 158.,  87., 202., 174., 137., 323.,
        77.,  42.,  38.,  80.,  79.,  99.,  83.,  86.,  79.,  48.,  53.,
        80., 110.,  93.,  70.,  71., 109.,  62.,  59.,  80.,  92., 143.,
        63., 131.,  61.,  70.,  67.,  81., 106.,  87.,  80.,  87.,  65.,
        56.,  58., 111., 237., 171., 114., 113., 131., 168., 127., 176.,
       137., 116.,  90., 114., 132., 156., 130., 134., 115., 109.,  88.,
       122., 148., 143., 100., 124.,  74., 116.,  88., 125., 160., 149.,
       167., 118., 148., 135., 102., 157., 124., 168., 174., 122., 180.,
       115., 132., 172., 110., 142., 118., 167., 112.,  83., 102.,  84.,
       164., 138., 175., 175., 144., 173., 123., 143.,  75.,  62.,  54.,
        81.,  62.,  99., 130., 101.,  56.,  64., 126., 174.,  82.,  81.,
        87., 115.,  54.,  91.,  80.,  90., 112., 123.,  87.,  87.,  52.,
        96.,  58.,  73.,  66.,  86.,  77., 116., 195., 147., 188., 165.,
       104., 112., 235., 189., 143., 143., 204., 221., 199., 136., 185.,
       191., 217., 170., 162., 267., 287., 240., 238., 232., 177., 168.,
       226., 197., 131., 217., 159., 178., 236., 170., 161., 181., 111.,
       150., 164., 241., 188., 183., 100., 184.,  92., 109., 124., 229.,
       156.,  83., 109., 108., 143.,  71., 186., 159., 126., 175., 112.,
        98., 141., 145.,  45.,  80.,  95.,  86.,  78., 130., 143.,  92.,
       117., 183.,  94., 113., 128., 154., 162., 112., 152., 255., 164.,
       162., 197., 252., 107., 100., 141., 104., 131., 250., 122.,  69.,
       109., 129., 269., 297., 152., 130., 104., 121., 118., 164.,  67.,
       170., 102., 143.,  95., 118., 167., 231., 252., 106.,  83., 107.,
       127., 190., 249., 155., 141.,  85., 136., 155., 210., 198., 143.,
       154., 143., 145., 188., 190., 235., 166., 242., 243., 166., 190.,
       313., 215., 128., 142., 251., 179., 144., 174., 219., 240., 169.,
       108., 143., 142., 170., 142., 133., 234., 153., 183., 140., 177.,
       206., 186., 211., 109., 122., 168., 172., 206., 229., 161., 174.,
       193., 148., 217., 230., 135., 222., 198., 145., 156., 164., 219.,
       198., 122., 126., 155., 220., 170., 152., 201., 217., 179., 146.,
       254., 246., 158., 227., 204., 150., 141., 150., 193.,  98., 104.,
       219.,  96., 236., 192., 155., 206., 163., 170., 180., 202., 153.,
       193., 152., 196., 186., 117., 139., 149., 243., 163., 174., 215.,
       221., 183., 114., 137., 117., 108., 315., 214., 188., 131., 198.,
       208., 153., 102., 202., 105., 141., 183., 196., 196., 227., 187.,
       210., 204., 210., 235., 188., 263., 161., 187., 180., 193., 155.,
       319., 201., 239., 182., 183., 208., 190., 185., 224., 180., 215.,
       188., 202., 164., 193., 208., 321., 177., 223., 238., 284., 209.,
       255., 231., 201., 180., 203., 208., 286., 244., 269., 171., 195.,
       180., 223., 177., 262., 227., 207., 192., 183., 237., 192., 230.,
       209., 312., 182., 233., 180., 242., 213., 155., 136., 283., 195.,
       140., 230., 225., 205., 195., 152., 168., 169., 127., 199., 218.,
       152., 131., 157., 196., 170., 158., 239., 145., 120., 149., 192.,
       128., 105.,  75., 129., 115., 156.,  82., 131.,  80., 118., 152.,
       237., 102., 151., 105., 109.,  94., 123., 136., 194., 157.,  97.,
       149., 132.,  87., 118., 117., 152., 126.,  96.,  96.,  73., 214.,
       180., 205., 220., 185., 121., 156., 216., 199., 190., 141., 199.,
```

```
126., 144., 144., 171., 256., 253., 203., 128., 153., 203., 103.,
121., 101., 103., 182., 161., 184., 172., 108., 179., 158., 146.,
179., 188., 177., 182., 235., 151., 106., 103., 129., 129., 170.,
152., 138., 114., 217., 167., 198., 216., 117., 184., 152., 146.,
110., 170., 183., 249., 201., 173., 261., 221., 190., 168., 229.,
143., 136., 133., 136., 167., 160., 174., 147., 116., 117., 181.,
253., 303., 192., 167., 113., 223., 176., 171., 191., 277., 212.,
132., 188., 164., 193., 187., 196., 158., 170., 201., 156., 211.,
191., 188., 185., 187., 208., 326., 191., 111., 252., 212., 151.,
149., 144., 187., 229., 183., 151., 206., 201., 218., 185., 250.,
198., 152., 144., 224., 155., 191., 176., 203., 257., 184., 157.,
180., 164., 198., 136., 175., 185., 180., 189., 210., 236., 179.,
209., 157., 135., 217., 182., 211., 264., 230., 211., 182., 179.,
307., 222., 251., 208., 160., 202., 205., 214., 168., 206., 297.,
189., 161., 196., 172., 233., 231., 195., 250., 167., 127., 250.,
205., 215., 260., 222., 181., 155., 249., 209., 223., 208., 227.,
255., 235., 240., 193., 225., 228., 159., 180., 197., 189., 168.,
249., 189., 146., 218., 125., 160., 216., 205., 199., 287., 191.,
185.])
```

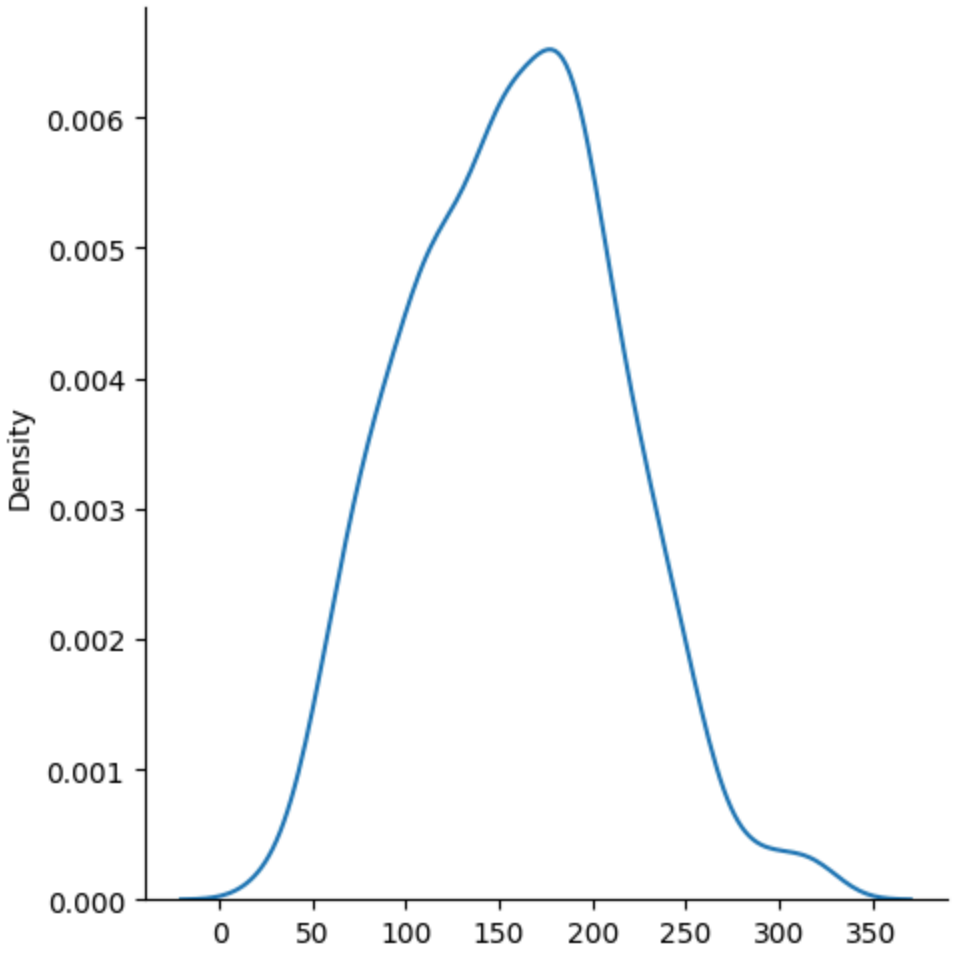
In [83]:
```
sns.distplot(guest_arrival_series.values)
```

Out[83]: <Axes: ylabel='Density'>



In [85]:
```
sns.displot(guest_arrival_series.values , kind = 'kde')
```

Out[85]: <seaborn.axisgrid.FacetGrid at 0x27a5ab64980>


Density
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
50
100
150
200
250
300
350

In [ ]: