Importing Libraries

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

Loading the dataset

```
df = pd.read_csv('dataset.csv')
```

EDA Exploratory Data Analysis and Data Cleaning

```
df.head()
```

```
{"type":"dataframe","variable_name":"df"}
```

```
df.tail()
```

```
{"type":"dataframe"}
```

```
df.shape
```

```
(119390, 32)
```

```
df.columns
```

```
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')
```

```
df.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
```

```
# Convert 'reservation_status_date' to datetime
df['reservation_status_date'] =
pd.to_datetime(df['reservation_status_date'], format='%d/%m/%Y')

df.describe(include = 'object')
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"hotel\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 4,\n
\"samples\": [\n          2,\n          \"79330\",\n
\"119390\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"arrival_date_month\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          12,\n          \"13877\",\n          \"119390\"\
n        ],\n        \"semantic_type\": \"\",\n
```

\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"meal\",\n   \"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 4,\n   \"samples\": [\n   5,\n
\"92310\",\n   \"119390\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"country\",\n   \"properties\":
{\n   \"dtype\": \"string\",\n   \"num_unique_values\": 4,\n
\"samples\": [\n   177,\n   \"48590\",\n
\"118902\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"market_segment\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   8,\n   \"56477\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"distribution_channel\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   5,\n   \"97870\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"reserved_room_type\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   10,\n   \"85994\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"assigned_room_type\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   12,\n   \"74053\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"deposit_type\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   3,\n   \"104641\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"customer_type\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   4,\n   \"89613\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"reservation_status\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 4,\n   \"samples\":
[\n   3,\n   \"75166\",\n   \"119390\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
for col in df.describe(include = 'object').columns:
    print(col)
    print(df[col].unique())
    print('-'*50)
```

```
hotel
['Resort Hotel' 'City Hotel']
--------------------------------------------------
arrival_date_month
['July' 'August' 'September' 'October' 'November' 'December' 'January'
 'February' 'March' 'April' 'May' 'June']
--------------------------------------------------
meal
['BB' 'FB' 'HB' 'SC' 'Undefined']
--------------------------------------------------
country
['PRT' 'GBR' 'USA' 'ESP' 'IRL' 'FRA' nan 'ROU' 'NOR' 'OMN' 'ARG' 'POL'
 'DEU' 'BEL' 'CHE' 'CN' 'GRC' 'ITA' 'NLD' 'DNK' 'RUS' 'SWE' 'AUS'
'EST'
 'CZE' 'BRA' 'FIN' 'MOZ' 'BWA' 'LUX' 'SVN' 'ALB' 'IND' 'CHN' 'MEX'
'MAR'
 'UKR' 'SMR' 'LVA' 'PRI' 'SRB' 'CHL' 'AUT' 'BLR' 'LTU' 'TUR' 'ZAF'
'AGO'
 'ISR' 'CYM' 'ZMB' 'CPV' 'ZWE' 'DZA' 'KOR' 'CRI' 'HUN' 'ARE' 'TUN'
'JAM'
 'HRV' 'HKG' 'IRN' 'GEO' 'AND' 'GIB' 'URY' 'JEY' 'CAF' 'CYP' 'COL'
'GGY'
 'KWT' 'NGA' 'MDV' 'VEN' 'SVK' 'FJI' 'KAZ' 'PAK' 'IDN' 'LBN' 'PHL'
'SEN'
 'SYC' 'AZE' 'BHR' 'NZL' 'THA' 'DOM' 'MKD' 'MYS' 'ARM' 'JPN' 'LKA'
'CUB'
 'CMR' 'BIH' 'MUS' 'COM' 'SUR' 'UGA' 'BGR' 'CIV' 'JOR' 'SYR' 'SGP'
'BDI'
 'SAU' 'VNM' 'PLW' 'QAT' 'EGY' 'PER' 'MLT' 'MWI' 'ECU' 'MDG' 'ISL'
'UZB'
 'NPL' 'BHS' 'MAC' 'TGO' 'TWN' 'DJI' 'STP' 'KNA' 'ETH' 'IRQ' 'HND'
'RWA'
 'KHM' 'MCO' 'BGD' 'IMN' 'TJK' 'NIC' 'BEN' 'VGB' 'TZA' 'GAB' 'GHA'
'TMP'
 'GLP' 'KEN' 'LIE' 'GNB' 'MNE' 'UMI' 'MYT' 'FRO' 'MMR' 'PAN' 'BFA'
'LBY'
 'MLI' 'NAM' 'BOL' 'PRY' 'BRB' 'ABW' 'AIA' 'SLV' 'DMA' 'PYF' 'GUY'
'LCA'
 'ATA' 'GTM' 'ASM' 'MRT' 'NCL' 'KIR' 'SDN' 'ATF' 'SLE' 'LAO']
--------------------------------------------------
market_segment
['Direct' 'Corporate' 'Online TA' 'Offline TA/TO' 'Complementary'
'Groups'
 'Undefined' 'Aviation']
--------------------------------------------------
distribution_channel
['Direct' 'Corporate' 'TA/TO' 'Undefined' 'GDS']
--------------------------------------------------
reserved_room_type
['C' 'A' 'D' 'E' 'G' 'F' 'H' 'L' 'P' 'B']
```

```
--------------------------------------------------
assigned_room_type
['C' 'A' 'D' 'E' 'G' 'F' 'I' 'B' 'H' 'P' 'L' 'K']
--------------------------------------------------
deposit_type
['No Deposit' 'Refundable' 'Non Refund']
--------------------------------------------------
customer_type
['Transient' 'Contract' 'Transient-Party' 'Group']
--------------------------------------------------
reservation_status
['Check-Out' 'Canceled' 'No-Show']
--------------------------------------------------
```

```
df.isnull().sum()
```

```
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                            4
babies                              0
meal                                0
country                           488
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
agent                           16340
company                        112593
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
```

```
df.dropna().isnull().sum()
```

```
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
```

```
df.describe()
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"is_canceled\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 42210.59518981826,\n
\"min\": 0.0,\n        \"max\": 119390.0,\n
\"num_unique_values\": 5,\n        \"samples\": [\n
0.37041628277075134,\n          0.48291822659316763,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"lead_time\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42151.05353185858,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 8,\n        \"samples\": [\n
104.01141636652986,\n          160.0,\n          119390.0\n        ],\
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"arrival_date_year\",\n
```

```
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
41605.6829984143,\n        \"min\": 0.7074759445202401,\n
\"max\": 119390.0,\n        \"num_unique_values\": 6,\n
\"samples\": [\n          119390.0,\n          2016.156554150264,\n
0.7074759445202401\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"arrival_date_week_number\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 42201.81402111843,\n
\"min\": 1.0,\n        \"max\": 119390.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
27.16517296255968,\n          38.0,\n          119390.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"arrival_date_day_of_month\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42205.50879379506,\n        \"min\": 1.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 8,\n        \"samples\": [\n
15.798241058715135,\n          23.0,\n          119390.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"stays_in_weekend_nights\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42209.53133083881,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 7,\n        \"samples\": [\n
119390.0,\n          0.9275986263506156,\n          19.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"stays_in_week_nights\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42207.69159751826,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 8,\n        \"samples\": [\n
2.500301532791691,\n          3.0,\n          119390.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"adults\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
42207.539491149604,\n        \"min\": 0.0,\n        \"max\":
119390.0,\n        \"num_unique_values\": 6,\n        \"samples\": [\n
119390.0,\n          1.8564033838679956,\n
0.5792609988330665\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"children\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 42208.79477831908,\n        \"min\":
0.0,\n        \"max\": 119386.0,\n        \"num_unique_values\": 5,\n
\"samples\": [\n          0.10388990333874994,\n
0.398561444786706,\n          0.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"babies\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
42210.22904846642,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.007948739425412514,\n          0.09743619130130332,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
```

```
}\n    },\n    {\n      \"column\": \"is_repeated_guest\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42210.678307305345,\n        \"min\": 0.0,\n        \"max\":
119390.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.03191222045397437,\n          0.17576714541079036,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"previous_cancellations\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42209.38002891464,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.08711784906608594,\n          0.8443363841518928,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\":
\"previous_bookings_not_canceled\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 42207.02766341039,\n
\"min\": 0.0,\n        \"max\": 119390.0,\n
\"num_unique_values\": 5,\n        \"samples\": [\n
0.13709690928888515,\n          1.4974368477089461,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"booking_changes\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42209.63515931599,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.22112404724013737,\n          0.6523055726747069,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"days_in_waiting_list\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42190.20344235136,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
2.321149174972778,\n          17.594720878752014,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"adr\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 41957.436659586354,\
n        \"min\": -6.38,\n        \"max\": 119390.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
101.83112153446686,\n          126.0,\n          119390.0\n        ],\
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"required_car_parking_spaces\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42210.319786938184,\n        \"min\": 0.0,\n        \"max\":
119390.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.06251779881062065,\n          0.24529114746719846,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"total_of_special_requests\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
42210.36738912644,\n        \"min\": 0.0,\n        \"max\": 119390.0,\
n        \"num_unique_values\": 6,\n        \"samples\": [\n
119390.0,\n          0.5713627607002262,\n
0.7927984228089581\n        ],\n        \"semantic_type\": \"\",\n
```

\"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"reservation_status_date\",\n   \"properties\": {\n \"dtype\": \"date\",\n   \"min\": \"1970-01-01 00:00:00.000119390\",\n   \"max\": \"2017-09-14 00:00:00\",\n \"num_unique_values\": 7,\n   \"samples\": [\n \"119390\",\n   \"2016-07-30 00:24:47.883407104\",\n \"2017-02-08 00:00:00\"\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
df = df[df['adr']<5000]

df.describe()
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 8,\n \"fields\": [\n {\n   \"column\": \"is_canceled\",\n   \"properties\": {\n \"dtype\": \"number\",\n   \"std\": 40869.56721485206,\n \"min\": 0.0,\n   \"max\": 115597.0,\n \"num_unique_values\": 5,\n   \"samples\": [\n 0.36986253968528593,\n   0.48276936275999693,\n   0.0\n],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n }\n   },\n   {\n   \"column\": \"lead_time\",\n \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 40809.80185564527,\n   \"min\": 0.0,\n   \"max\": 115597.0,\n   \"num_unique_values\": 8,\n   \"samples\": [\n 105.01959393409864,\n   162.0,\n   115597.0\n   ],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n }\n   },\n   {\n   \"column\": \"arrival_date_year\",\n \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 40264.85470157284,\n   \"min\": 0.7069896291119789,\n \"max\": 115597.0,\n   \"num_unique_values\": 6,\n \"samples\": [\n   115597.0,\n   2016.144752891511,\n 0.7069896291119789\n   ],\n   \"semantic_type\": \"\",\n \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"arrival_date_week_number\",\n   \"properties\": {\n \"dtype\": \"number\",\n   \"std\": 40860.83523719635,\n \"min\": 1.0,\n   \"max\": 115597.0,\n \"num_unique_values\": 8,\n   \"samples\": [\n 27.045788385511734,\n   38.0,\n   115597.0\n   ],\n \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"arrival_date_day_of_month\",\n \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 40864.48210516812,\n   \"min\": 1.0,\n   \"max\": 115597.0,\n   \"num_unique_values\": 8,\n   \"samples\": [\n 15.770089189165809,\n   23.0,\n   115597.0\n   ],\n \"semantic_type\": \"\",\n   \"description\": \"\"\n   }\n   },\n   {\n   \"column\": \"stays_in_weekend_nights\",\n \"properties\": {\n   \"dtype\": \"number\",\n   \"std\": 40868.50372950941,\n   \"min\": 0.0,\n   \"max\": 115597.0,\n   \"num_unique_values\": 7,\n   \"samples\": [\n

```
115597.0,\n          0.9207678399958477,\n          19.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"stays_in_week_nights\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
40866.66477184999,\n        \"min\": 0.0,\n        \"max\": 115597.0,\
n        \"num_unique_values\": 8,\n        \"samples\": [\n
2.4814657819839616,\n          3.0,\n          115597.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"adults\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
40866.512138446815,\n        \"min\": 0.0,\n        \"max\":
115597.0,\n        \"num_unique_values\": 6,\n        \"samples\": [\n
115597.0,\n          1.8460340666280266,\n
0.5793226026996661\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"children\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 40867.77029884725,\n        \"min\":
0.0,\n        \"max\": 115593.0,\n        \"num_unique_values\": 5,\n
\"samples\": [\n          0.08267801683492945,\n
0.3499607735485756,\n          0.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"babies\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
40869.20118364985,\n        \"min\": 0.0,\n        \"max\": 115597.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.0074915438981980505,\n          0.09510277181586718,\n          0.0\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"is_repeated_guest\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 40869.65018514639,\n        \"min\":
0.0,\n        \"max\": 115597.0,\n        \"num_unique_values\": 5,\n
\"samples\": [\n          0.032518144934557125,\n
0.1773724537392609,\n          0.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"previous_cancellations\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
40868.35123214405,\n        \"min\": 0.0,\n        \"max\": 115597.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.08980336859953113,\n          0.857844399643005,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\":
\"previous_bookings_not_canceled\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 40865.9985148412,\n
\"min\": 0.0,\n        \"max\": 115597.0,\n
\"num_unique_values\": 5,\n        \"samples\": [\n
0.1408254539477668,\n          1.521073870276689,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"booking_changes\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
```

```
40868.60763556607,\n        \"min\": 0.0,\n        \"max\": 115597.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.21700390148533266,\n          0.6472000688286631,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"days_in_waiting_list\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
40849.165604258305,\n        \"min\": 0.0,\n        \"max\":
115597.0,\n        \"num_unique_values\": 5,\n        \"samples\": [\n
2.3876398176423264,\n          17.863820777442267,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"adr\",\n      \"properties\": {\
n        \"dtype\": \"number\",\n        \"std\": 40838.2599250576,\n
\"min\": -6.38,\n        \"max\": 115597.0,\n
\"num_unique_values\": 8,\n        \"samples\": [\n
97.03664299246519,\n          121.37,\n          115597.0\n        ],\
n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"required_car_parking_spaces\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
40869.29218488849,\n        \"min\": 0.0,\n        \"max\": 115597.0,\
n        \"num_unique_values\": 5,\n        \"samples\": [\n
0.059672828879642204,\n          0.2401089616347926,\n          0.0\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"total_of_special_requests\",\n
\"properties\": {\n        \"dtype\": \"number\",\n        \"std\":
40869.340246054584,\n        \"min\": 0.0,\n        \"max\":
115597.0,\n        \"num_unique_values\": 6,\n        \"samples\": [\n
115597.0,\n          0.5616754760071628,\n
0.7853326807198735\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"reservation_status_date\",\n      \"properties\": {\n
\"dtype\": \"date\",\n        \"min\": \"1970-01-01
00:00:00.000115597\",\n        \"max\": \"2017-09-14 00:00:00\",\n
\"num_unique_values\": 7,\n        \"samples\": [\n
\"115597\",\n          \"2016-07-24 13:40:20.128550144\",\n
\"2017-02-02 00:00:00\"\n        ],\n        \"semantic_type\": \"\",\
n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe"}
```

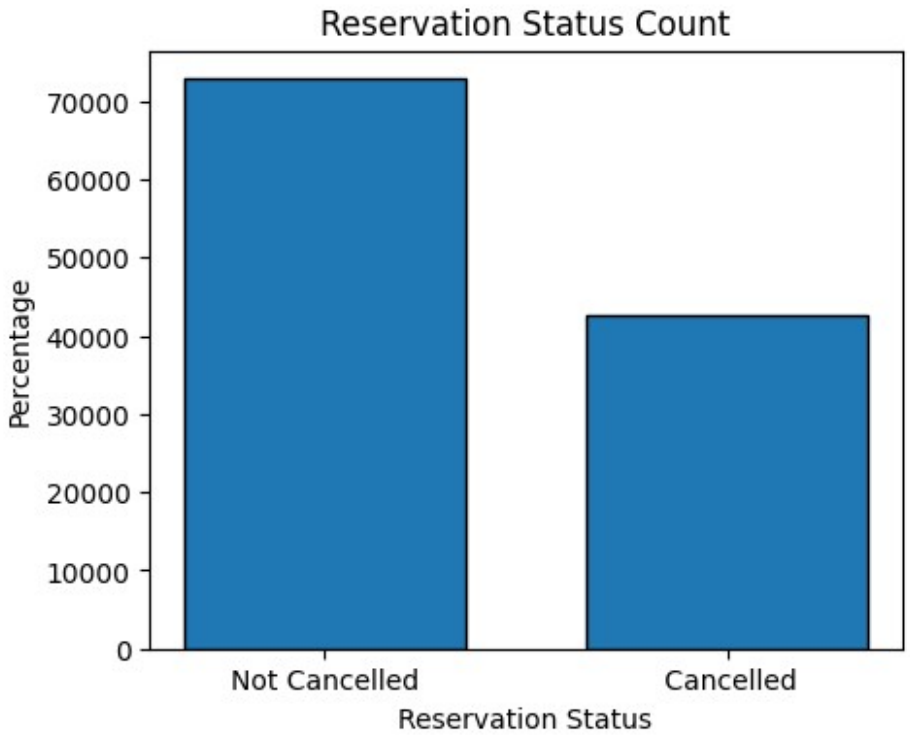

Data Analysis and Visualizations (Analysis And Findings)

```
cancelled_perc = df['is_canceled'].value_counts(normalize = True)
print(cancelled_perc)

plt.figure(figsize = (5,4))
plt.title('Reservation Status Count')
plt.bar(['Not Cancelled','
Cancelled'],df['is_canceled'].value_counts(),edgecolor = 'k',width =
0.7)
plt.xlabel('Reservation Status')
```

```
plt.ylabel('Percentage')
plt.show()
```

```
is_canceled
0    0.630137
1    0.369863
Name: proportion, dtype: float64
```



The accompanying bar graph shows the percentage of reservations that are canceled and those that are not. It is obvious that there are still a significant number of reservations that have not been canceled.There are still 37% of clients who canceled their reservation,which has a significant impact on the hotel's earning.

```
plt.figure(figsize = (8,4))
ax1=sns.countplot(x='hotel',hue='is_canceled',data=df,palette='Blues')
legend_labels,_ = ax1.get_legend_handles_labels()
plt.title('Reservation Status in different Hotels')
plt.xlabel('Hotel')
plt.ylabel('No of Reservations')
plt.legend(['Not Canceled', 'Canceled'])
plt.show()
```

In comparison to resort hotels,city hotels have more bookings.It's possible that resort hotel's are more expensive than thse in cities.

```
resort_hotel = df[df['hotel']=='Resort Hotel']
resort_hotel['is_canceled'].value_counts(normalize = True)
```

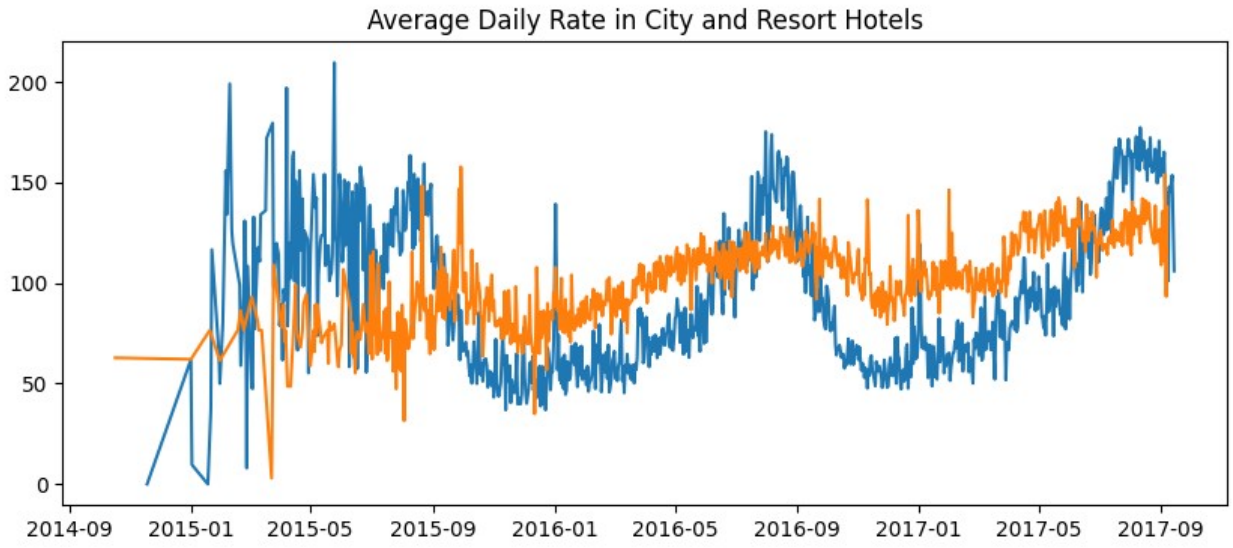
```
is_canceled
0    0.729877
1    0.270123
Name: proportion, dtype: float64
```

```
city_hotel = df[df['hotel']=='City Hotel']
city_hotel['is_canceled'].value_counts(normalize = True)
```

```
is_canceled
0    0.582137
1    0.417863
Name: proportion, dtype: float64
```

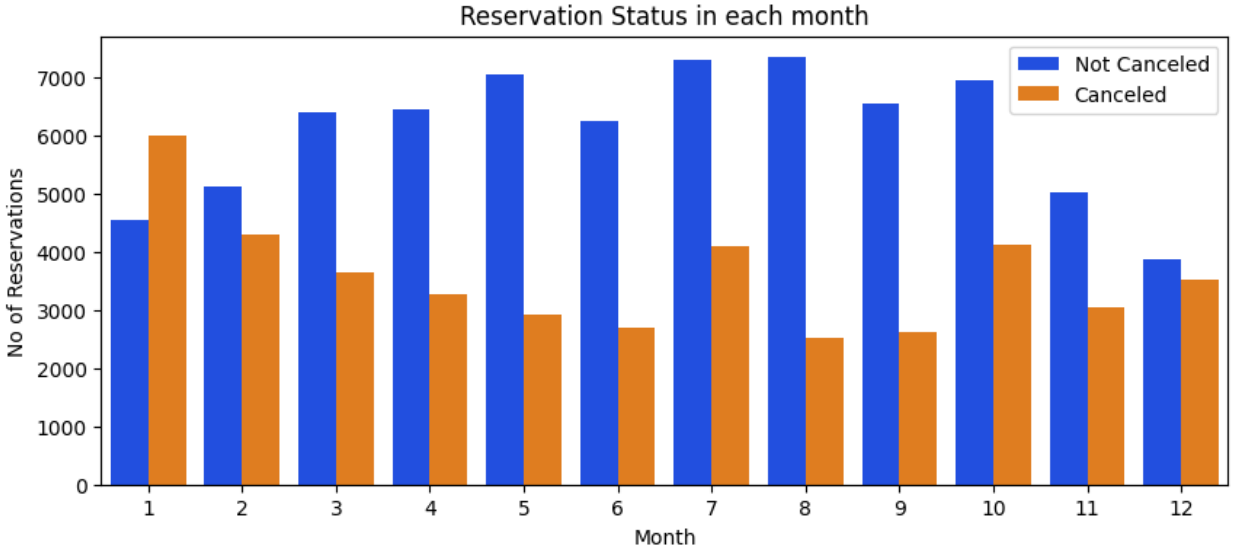
```
resort_hotel = resort_hotel.groupby('reservation_status_date')
['adr'].mean()
city_hotel = city_hotel.groupby('reservation_status_date')
['adr'].mean()

plt.figure(figsize = (10,4))
plt.title('Average Daily Rate in City and Resort Hotels')
plt.plot(resort_hotel.index, resort_hotel,label = 'Resort Hotel')
plt.plot(city_hotel.index, city_hotel,label = 'City Hotel')
plt.show()
```



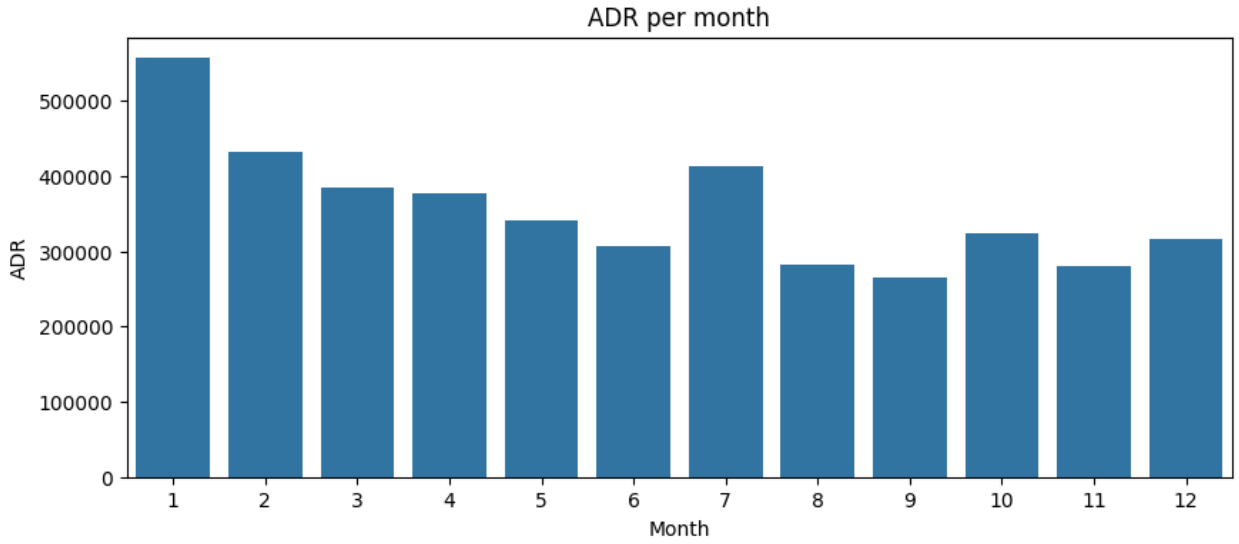
The line graph above shows that, on certain days, the average daily rate for a city hotel is less than that of a resort hotel, and on other days,it is even less. It goes without saying that weekends and hoildays may see a rise in resort hotel rates.

```
df['month'] = df['reservation_status_date'].dt.month
plt.figure(figsize=(10,4))
ax1 =
sns.countplot(x='month',hue='is_canceled',data=df,palette='bright')
legend_labels,_ = ax1.get_legend_handles_labels()
ax1.legend(bbox_to_anchor=(1,1))
plt.title('Reservation Status in each month')
plt.xlabel('Month')
plt.ylabel('No of Reservations')
plt.legend(['Not Canceled', 'Canceled'])
plt.show()
```



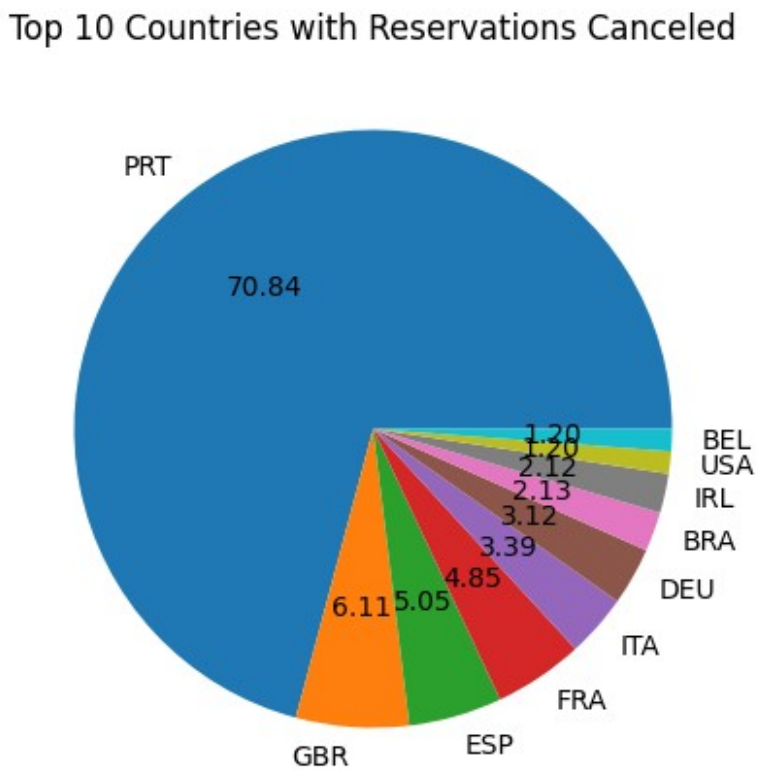
We have developed the grouped bar graph to analyze the months with the highest and lowest reservation levels according to reservations and the number of canceled reservations are largest in the month of august whereas January is the moth with the most canceled reservations.

```
plt.figure(figsize = (10,4))
plt.title('ADR per month')
sns.barplot(x='month', y='adr', data=df[df['is_canceled'] ==
1].groupby('month')[['adr']].sum().reset_index())
plt.xlabel('Month')
plt.ylabel('ADR')
plt.show()
```



This bar graph demonstrates that cancellations are most common when prices are greatest and are least common when they are lowest.Therefore,thecost of the accommodation is solely responsible for the cancellation.

```
cancelled_data = df[df['is_canceled'] == 1]
top_10_country = cancelled_data['country'].value_counts()[:10]
plt.figure(figsize = (10,5))
plt.title('Top 10 Countries with Reservations Canceled')
plt.pie(top_10_country, labels = top_10_country.index, autopct =
'%.2f')
plt.show()
```



Now, let's see which country has the highest reservation canceled. The top country is Portugal with the highest number of cancellations.

Let's check the area from where guest are visiting the hotels and making reservations.Is it coming from Direct, Group,Online,Offline Travel Agents? Around 46% of the clients come from online travel agencies,whereas 27% come from groups. Only 4% of clients book hotels directly by visiting them and making reservations.

```
df['market_segment'].value_counts()
```

```
market_segment
Online TA        53786
Offline TA/TO    24111
Groups           19712
Direct           11717
Corporate         5289
Complementary      743
Aviation           237
Undefined            2
Name: count, dtype: int64
```

```
df['market_segment'].value_counts(normalize=True)
```

```
market_segment
Online TA        0.465289
```

```
Offline TA/TO    0.208578
Groups           0.170523
Direct           0.101361
Corporate        0.045754
Complementary    0.006428
Aviation         0.002050
Undefined        0.000017
Name: proportion, dtype: float64
```

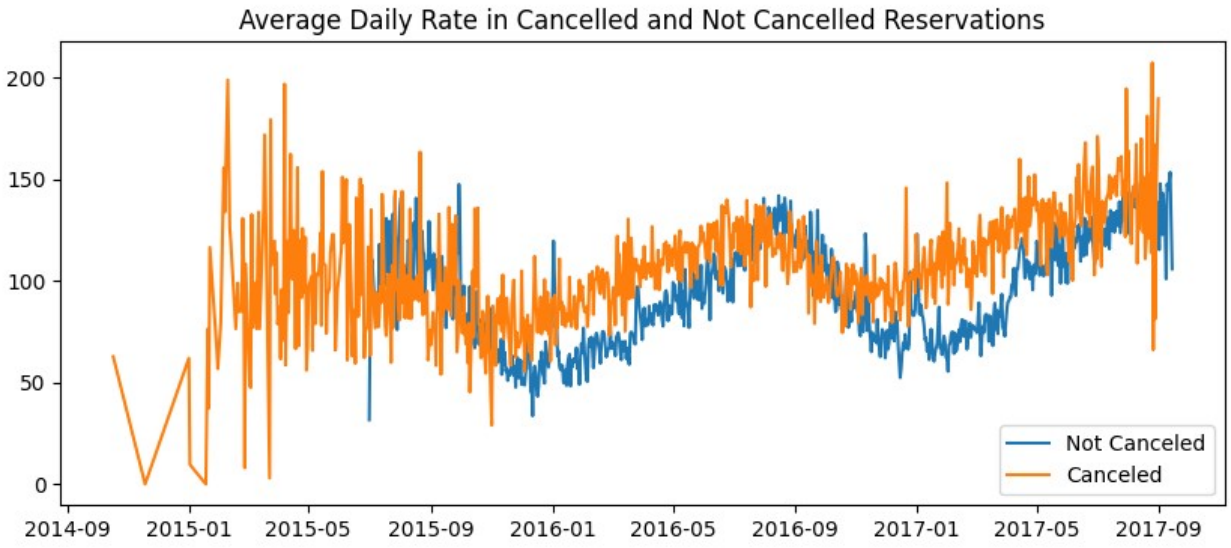
```
cancelled_data['market_segment'].value_counts(normalize=True)
```

```
market_segment
Online TA        0.456321
Groups           0.282049
Offline TA/TO    0.193919
Direct           0.041024
Corporate        0.023155
Complementary    0.002269
Aviation         0.001216
Undefined        0.000047
Name: proportion, dtype: float64
```

```
cancelled_df_adr = cancelled_data.groupby('reservation_status_date')
['adr'].mean()
cancelled_df_adr = cancelled_df_adr.reset_index()
cancelled_df_adr.sort_values('reservation_status_date', inplace =
True)

not_cancelled_df_adr = df[df['is_canceled'] == 0]
not_cancelled_df_adr =
not_cancelled_df_adr.groupby('reservation_status_date')['adr'].mean()
not_cancelled_df_adr = not_cancelled_df_adr.reset_index()
not_cancelled_df_adr.sort_values('reservation_status_date', inplace =
True)

plt.figure(figsize = (10,4))
plt.title('Average Daily Rate in Cancelled and Not Cancelled
Reservations')
plt.plot(not_cancelled_df_adr['reservation_status_date'],
not_cancelled_df_adr['adr'],label = 'Not Canceled')
plt.plot(cancelled_df_adr['reservation_status_date'],
cancelled_df_adr['adr'],label = 'Canceled')
plt.legend()
plt.show()
```



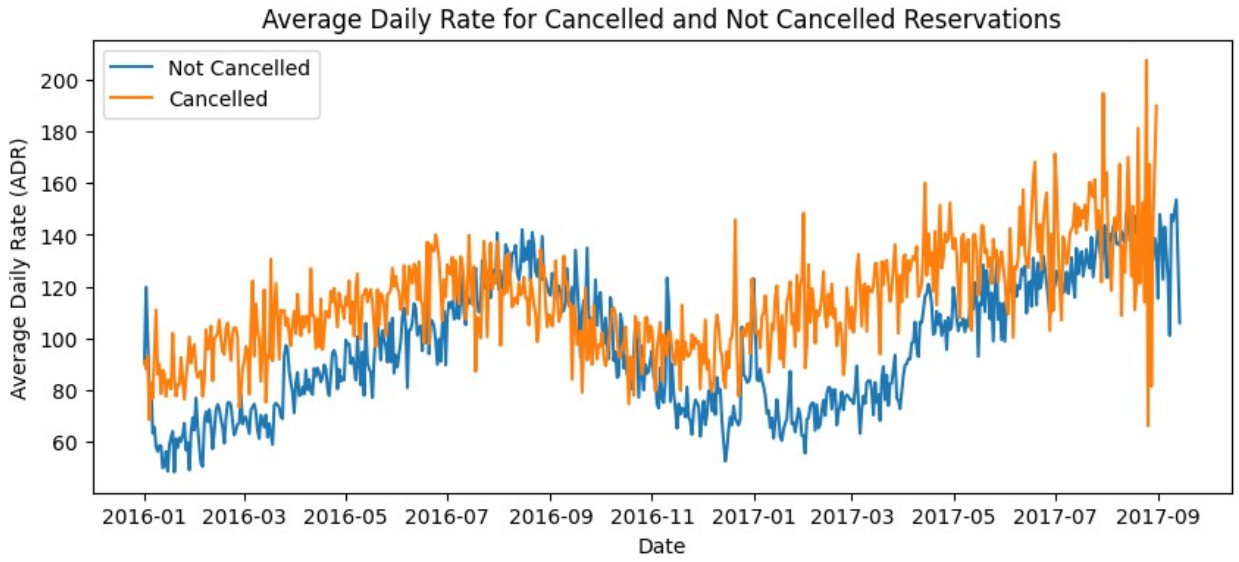

```
filtered_df = df[(df['reservation_status_date'] >= '2016-01-01') &
(df['reservation_status_date'] <= '2017-09-30')]

plt.figure(figsize=(10, 4))
plt.title('Average Daily Rate for Cancelled and Not Cancelled
Reservations')

# Filter data for the specified date range
cancelled_df_filtered =
cancelled_df_adr[(cancelled_df_adr['reservation_status_date'] >=
'2016-01-01') & (cancelled_df_adr['reservation_status_date'] <= '2017-
09-30')]
not_cancelled_df_filtered =
not_cancelled_df_adr[(not_cancelled_df_adr['reservation_status_date']
>= '2016-01-01') & (not_cancelled_df_adr['reservation_status_date'] <=
'2017-09-30')]

plt.plot(not_cancelled_df_filtered['reservation_status_date'],
not_cancelled_df_filtered['adr'], label='Not Cancelled')
plt.plot(cancelled_df_filtered['reservation_status_date'],
cancelled_df_filtered['adr'], label='Cancelled')

plt.xlabel('Date')
plt.ylabel('Average Daily Rate (ADR)')
plt.legend()
plt.show()
```



As seen in thhe graph, reservations are cancelled when the average daily rate is higher than when it is not canceled.It clearly proves all the above analysis,that the higher price leads to higher cancellation.

Suggestions

1.Cancellation rates rise as the price does. In order to prvent cancellations of reservations,hotel could work on their pricing strategies and try to lower the rates for specific hotels based on locations.They can also provide some discounts to the consumers.

2.As the ratio of the cancellation and not cancellation of the resort hotel in higher the resort hotel than the city hotels. So the hotels should provide a reasonable discount on the room prices on weekends or on hoildays.

3.In the month of January, hotels can start campaigns or making with a reasonable amount to increase their revenue as the cancellation is the highest in this month.

4.they can also increase the quality of their hotels and their services mainly in Portugal to reduce the cancellation rate.