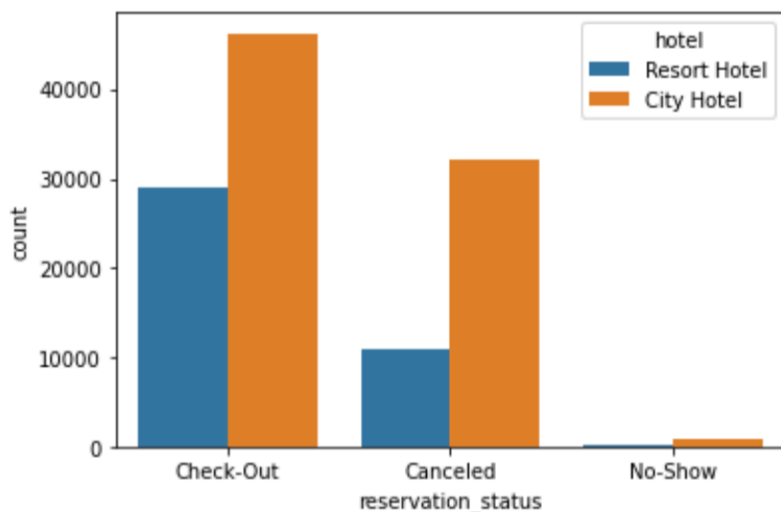


צרו איור המציג השוואה בין בשני סוגי המלונות מבחינת כמות ההזמנות בכל סטטוס

```
[76]: sns.countplot(x= 'reservation_status',hue = 'hotel',data = df) # hue - מה שנרצה להשוואת  
#x גורם היחס
```

```
[76]: <AxesSubplot:xlabel='reservation_status', ylabel='count'>
```



ענייננו להשוואת שני סוגי המלונות מבחינת כמות ההזמנות בכל סטטוס

צרו טבלה המציגה את סך הכל מספר המבוגרים בכל חודש בשני סוגי המלונות

```
[79]: df.pivot_table('adults', index = 'hotel', columns='arrival_date_month', aggfunc="count")
```

```
[79]:
```

arrival_date_month	April	August	December	February	January	July	June	March	May	November	October	September
hotel												
City Hotel	7480	8983	4132	4965	3736	8088	7894	6458	8232	4357	7605	7400
Resort Hotel	3609	4894	2648	3103	2193	4573	3045	3336	3559	2437	3555	3108

קולמנס מסדר לנו עמודות לפי הערכים הקיימים באותה עמודה

```
[90]: correlation_matrix = df[:].corr(method="pearson") # עובר על כול העמודות
correlation_matrix.style.background_gradient(cmap='coolwarm')
```

	is_canceled	lead_time	arrival_date_year	arrival_date_week_number	arrival_date_day_of_month	stays_in_weekend_nights
is_canceled	1.000000	0.293123	0.016660	0.008148	-0.006130	-0.001791
lead_time	0.293123	1.000000	0.040142	0.126871	0.002268	0.024765
arrival_date_year	0.016660	0.040142	1.000000	-0.540561	-0.000221	0.029218
arrival_date_week_number	0.008148	0.126871	-0.540561	1.000000	0.066809	-0.020904
arrival_date_day_of_month	-0.006130	0.002268	-0.000221	0.066809	1.000000	0.010613
stays_in_weekend_nights	-0.001791	0.085671	0.021497	0.018208	-0.016354	1.000000
stays_in_week_nights	0.024765	0.165799	0.030883	0.015558	-0.028174	0.000300
adults	0.060017	0.119519	0.029635	0.025909	-0.001566	0.000149
children	0.005048	-0.037622	0.054624	0.005518	0.014544	0.063457
babies	-0.032491	-0.020915	-0.013192	0.010395	-0.000230	-0.076760
is_repeated_guest	-0.084793	-0.124410	0.010341	-0.030131	-0.006145	0.022933
previous_cancellations	0.110133	0.086042	-0.119822	0.035501	-0.027011	0.075791
previous_bookings_not_canceled	-0.057358	-0.073548	0.029218	-0.020904	-0.000300	0.030245
booking_changes	-0.144381	0.000149	0.030872	0.005508	0.010613	0.008683
agent	-0.083114	-0.069741	0.063457	-0.031201	0.001487	
company	-0.020642	0.151464	0.259095	-0.076760	0.044858	
days_in_waiting_list	0.054186	0.170084	-0.056497	0.022933	0.022728	
adr	0.047557	-0.063077	0.197580	0.075791	0.030245	
required_car_parking_spaces	-0.195498	-0.116451	-0.013684	0.001920	0.008683	

```
type=object ,
```

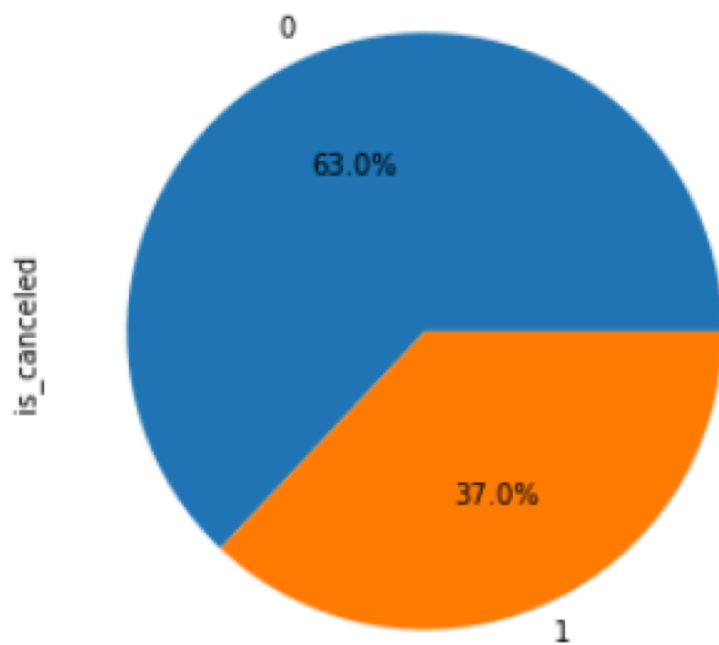
צרו מודל המנבא האם ההזמנה בוטלה, על סמך אפשרויות המופיעים ב קובץ הנלמד הציגו את הדיוק של המודל. האם הדיוק הזה הוא יותר טוב מדיוק מקרי? נמקו את התשובה שלכם

```
60]: features = ['stays_in_weekend_nights','previous_cancellations',
               'stays_in_week_nights','babies','is_repeated_guest',
               'previous_bookings_not_canceled','days_in_waiting_list',
               'lead_time']
X = df[features]
y = df['is_canceled'] #וואי יהיה מה שנרצה לנבאות
X_train, X_test, y_train, y_test = sk.model_selection.train_test_split(X,y,test_size=0.3,random_state=1)
model = sk.tree.DecisionTreeClassifier(random_state= 0)
model.fit(X_train,y_train) # (מודל מכניסים את טראין (את ערכי "האימון"
pred = model.predict(X_test)
sk.metrics.accuracy_score(pred,y_test)
```

```
60]: 0.760253510902644
```

תשובה: כל דיוק מעל 63% הוא יותר טוב מדיוק מקרי

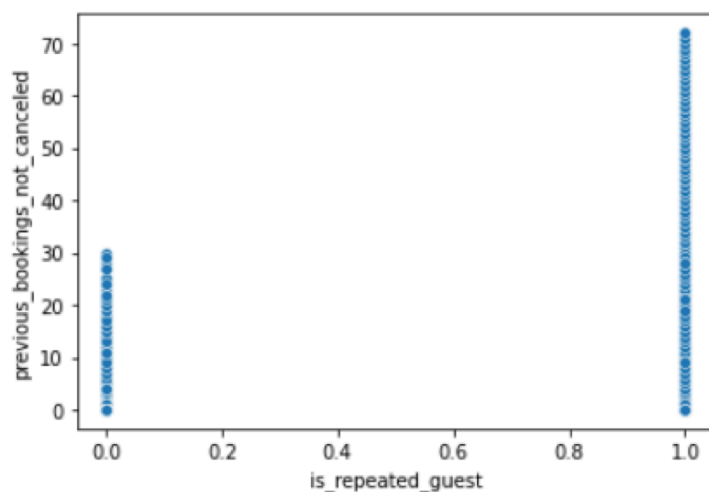
```
54]: plt.figure(1, figsize=(5,5))  
data['is_canceled'].value_counts().plot.pie(autopct="%1.1f%%")  
plt.show()
```



בדקו האם קיימות קורלציה בין ה features המופיעים ב test\_a.ipynb  
צרו scatterplot עבור הקורלציה החיובית השנייה הכי חזקה שמצאתם.

```
[25]: sns.scatterplot(x = 'is_repeated_guest', y = 'previous_bookings_not_canceled', data = data)
```

```
[25]: <AxesSubplot:xlabel='is_repeated_guest', ylabel='previous_bookings_not_canceled'>
```

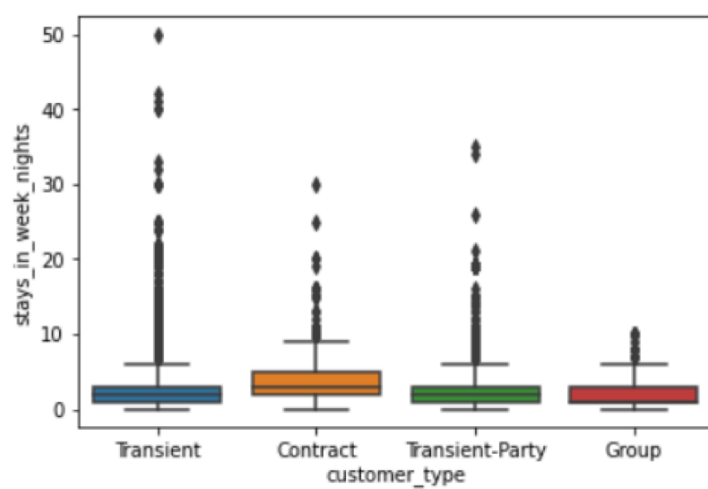


### שאלה 3 (15%)

צרו איור המציג את סוג הלקוחות (customer\_type) ומספר הלילות שישנו במלון (stays\_in\_week\_nights)

Q3

```
[12]: sns.boxplot(y='stays_in_week_nights', x='customer_type', data=data);
```



צרו טבלה המציגה את מספר הילדים (children) הממוצע בכל שנה (arrival\_date\_year) בשני סוגי המלונות (hotel)

## Q2

```
[9]: data.pivot_table('children',index='hotel', columns='arrival_date_year', aggfunc='mean')
```

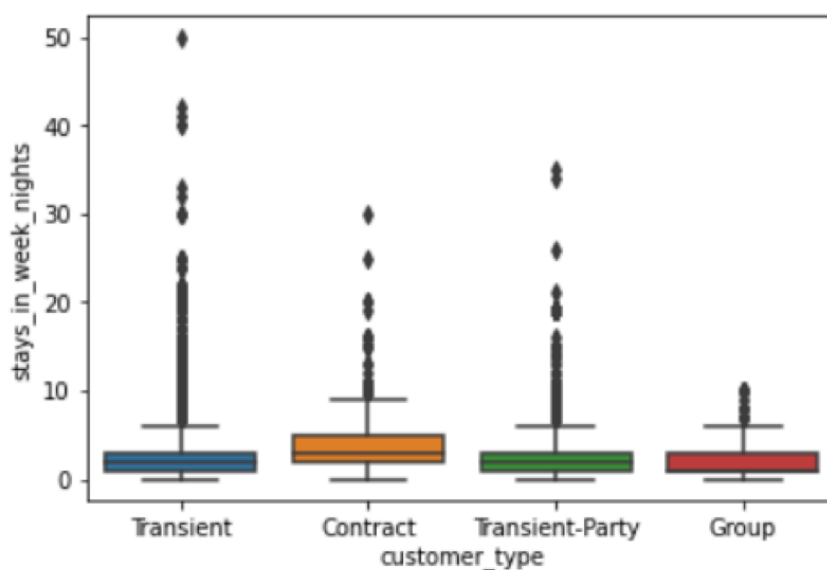
```
[9]: arrival_date_year    2015    2016    2017
      hotel
City Hotel    0.037432  0.100000  0.106224
Resort Hotel   0.101636  0.112781  0.168146
```

### שאלה 3 (15%)

צרו איור המציג את סוג הלקוחות (customer\_type) ומספר הלילות שישנו במלון (stays\_in\_week\_nights)

## Q3

```
[12]: sns.boxplot(y='stays_in_week_nights', x='customer_type', data=data);
```





צרו איור המציג השוואה בין שני סוגי המלונות (hotel) מבחינת כמות הביטולים (is\_canceled)

## Q1

```
[6]: sns.countplot(x="is_canceled", hue="hotel", data=data)
```

```
[6]: <AxesSubplot:xlabel='is_canceled', ylabel='count'>
```

