In [9]:

```python
# 1. Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:

```python
# 2. Load dataset
df = pd.read_csv(r"C:\Users\K KRISHNAVINAYAKA\Downloads\Hotel Bookings.csv")
```

In [3]:

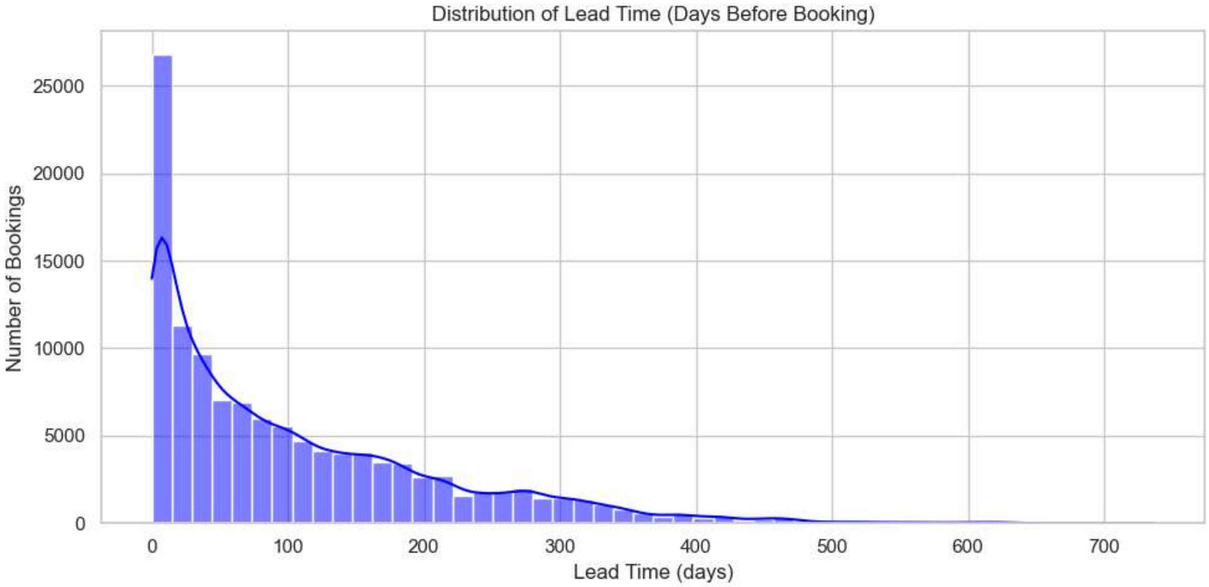
```python
# 3. Preprocessing
df['total_stay'] = df['stays_in_weekend_nights'] + df['stays_in_week_nights']
df['reservation_status_date'] = pd.to_datetime(df['reservation_status_date'])
```

In [4]:

```python
# Set style
sns.set(style="whitegrid")
```

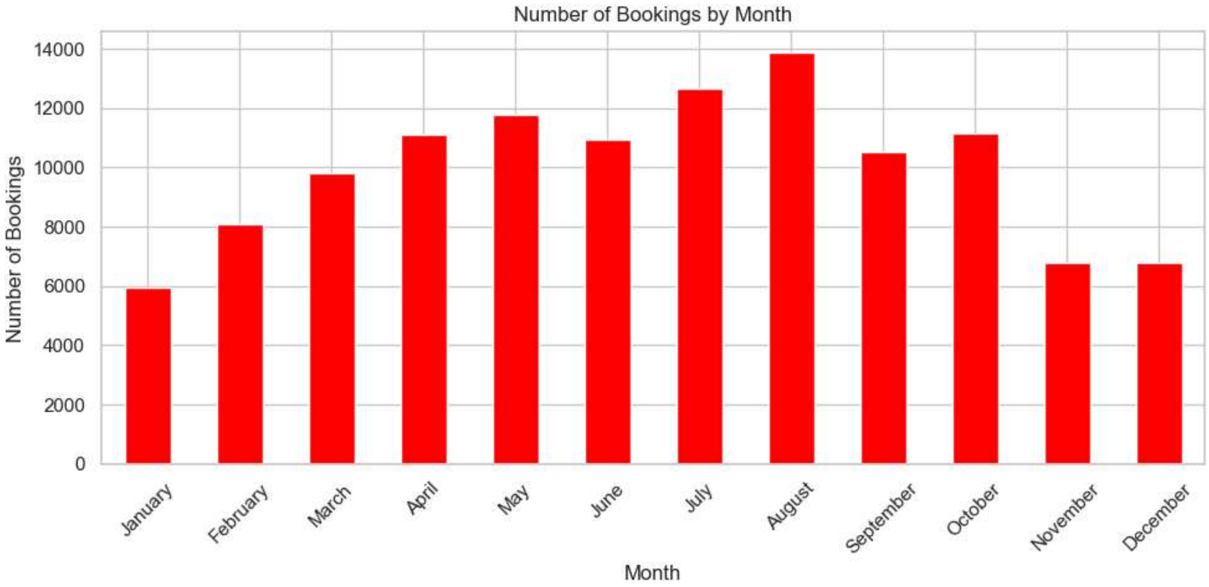
In [36]:

```python
# Section 1: Lead Time Analysis ---
plt.figure(figsize=(10, 5))
sns.histplot(df['lead_time'], bins=50, kde=True, color='blue')
plt.title('Distribution of Lead Time (Days Before Booking)')
plt.xlabel('Lead Time (days)')
plt.ylabel('Number of Bookings')
plt.tight_layout()
plt.show()
```



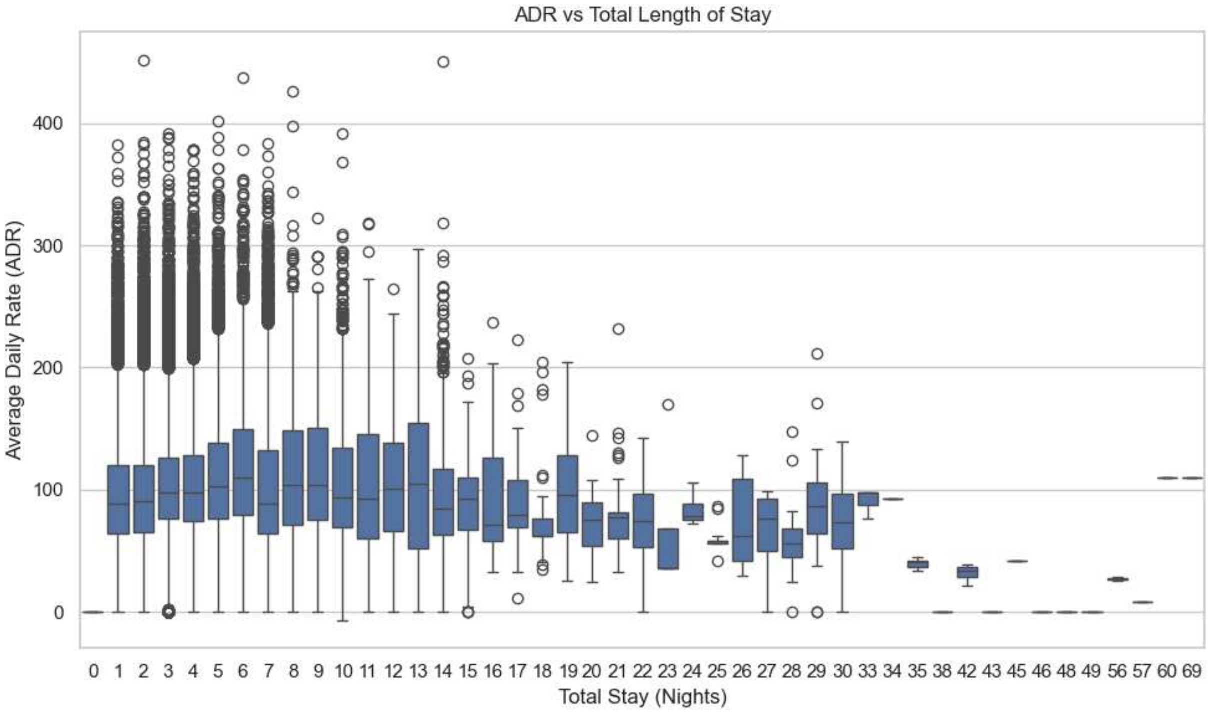
In [30]:

```python
# Section 2: Monthly Bookings ---
monthly_bookings = df['arrival_date_month'].value_counts().reindex([
    'January', 'February', 'March', 'April', 'May', 'June',
    'July', 'August', 'September', 'October', 'November', 'December'
])
monthly_bookings.plot(kind='bar', color='red', figsize=(10, 5))
plt.title('Number of Bookings by Month')
plt.xlabel('Month')
plt.ylabel('Number of Bookings')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



In [27]:

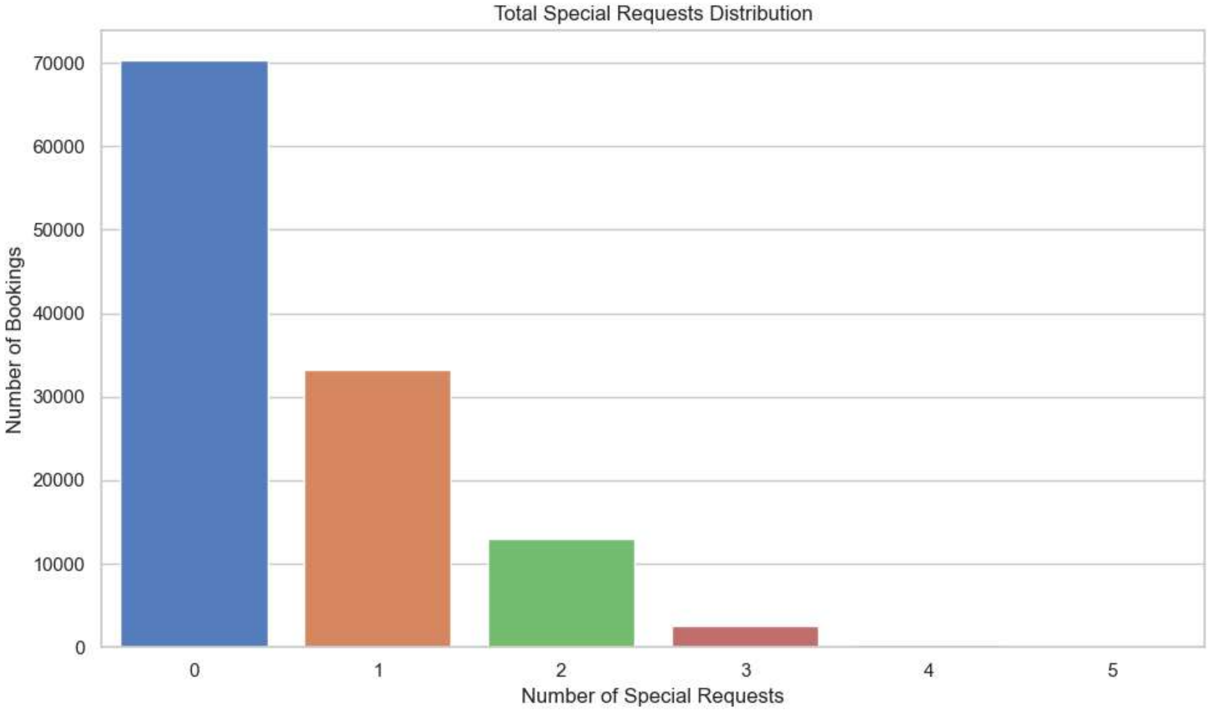
```
# Section 3: Stay Length vs ADR ---
filtered_df = df[df['adr'] < 500]
plt.figure(figsize=(10, 6))
sns.boxplot(x='total_stay', y='adr', data=filtered_df)
plt.title('ADR vs Total Length of Stay')
plt.xlabel('Total Stay (Nights)')
plt.ylabel('Average Daily Rate (ADR)')
plt.tight_layout()
plt.show()
```



In [26]:

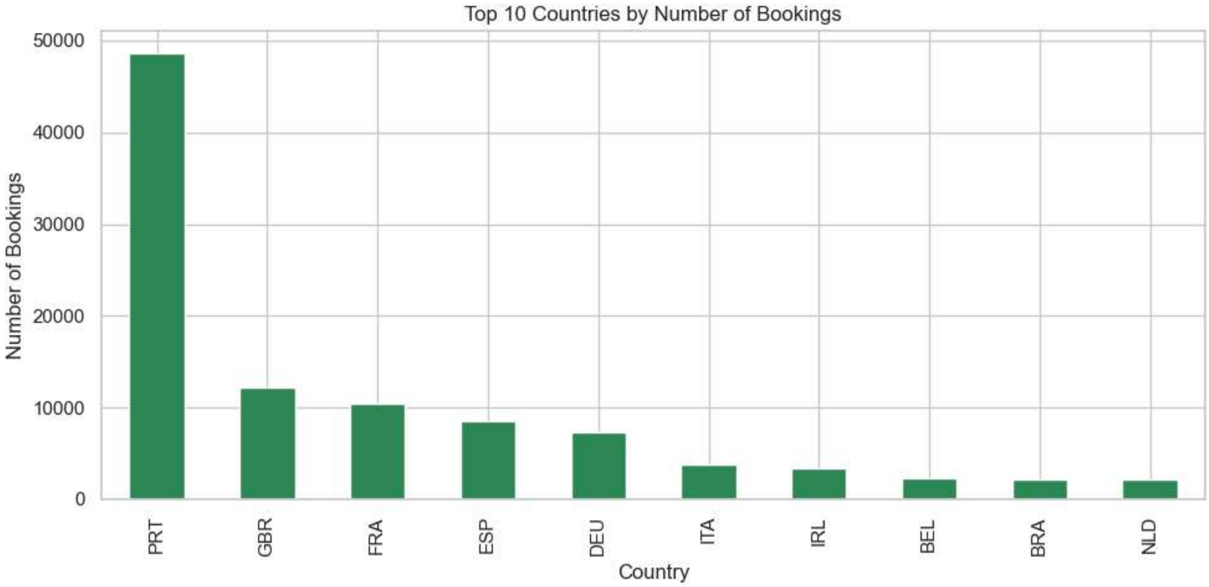
```
# Section 4: Special Requests Distribution ---
plt.figure(figsize=(10, 6))
```

```
sns.countplot(x='total_of_special_requests', data=df, hue='total_of_special_requests'
plt.title('Total Special Requests Distribution')
plt.xlabel('Number of Special Requests')
plt.ylabel('Number of Bookings')
plt.tight_layout()
plt.show()
```



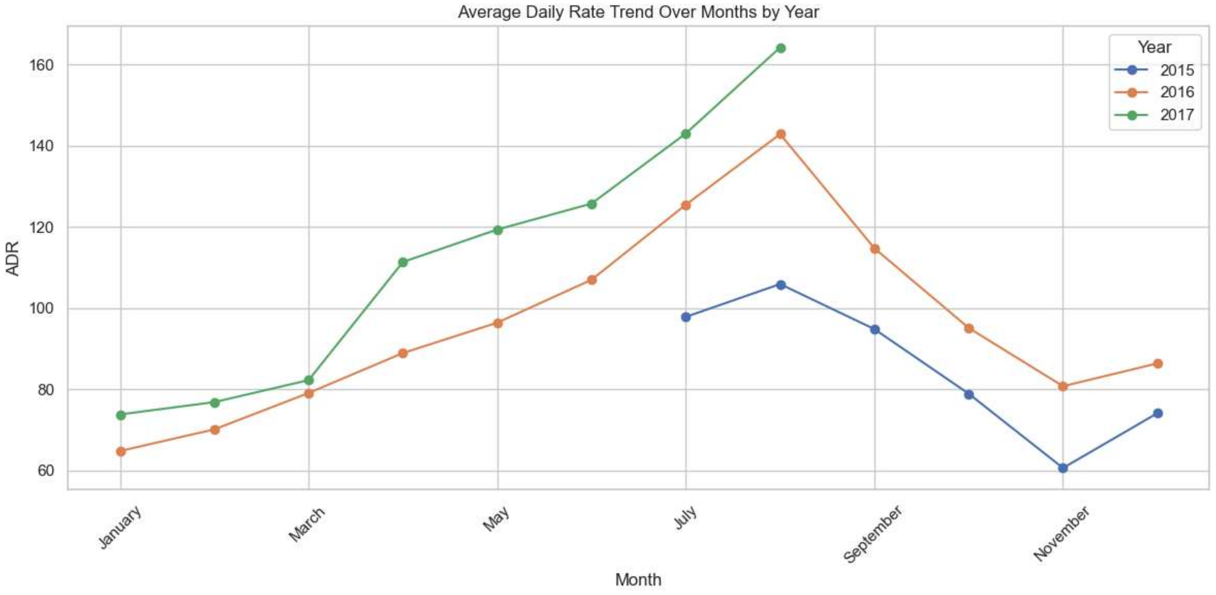
In [25]:

```
# Section 5: Bookings by Country ---
top_countries = df['country'].value_counts().head(10)
top_countries.plot(kind='bar', figsize=(10, 5), color='seagreen')
plt.title('Top 10 Countries by Number of Bookings')
plt.xlabel('Country')
plt.ylabel('Number of Bookings')
plt.tight_layout()
plt.show()
```



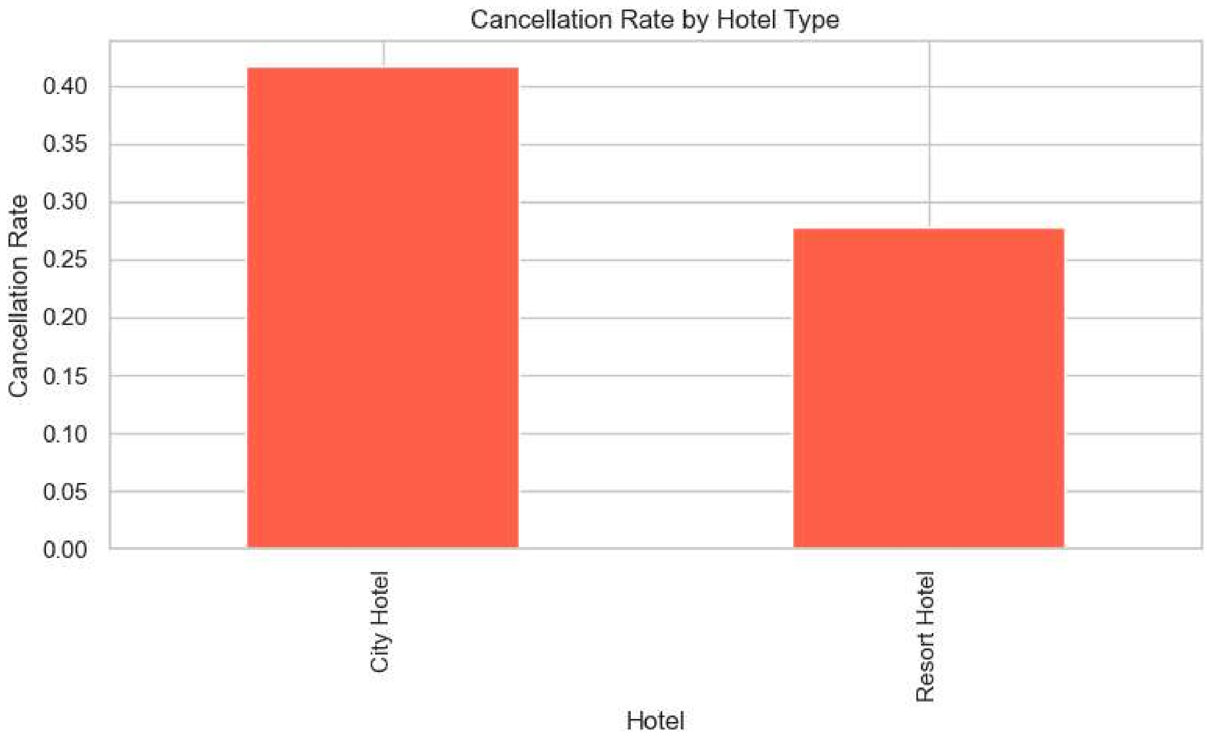
In [24]:

```
# Section 6: ADR Trends by Month and Year ---
monthly_adr = df.groupby(['arrival_date_year', 'arrival_date_month'])['adr'].mean().u
months_ordered = ['January', 'February', 'March', 'April', 'May', 'June',
                  'July', 'August', 'September', 'October', 'November', 'December']
monthly_adr = monthly_adr.loc[months_ordered]
monthly_adr.plot(figsize=(12, 6), marker='o')
plt.title('Average Daily Rate Trend Over Months by Year')
plt.ylabel('ADR')
plt.xlabel('Month')
plt.xticks(rotation=45)
plt.legend(title='Year')
plt.tight_layout()
plt.show()
```



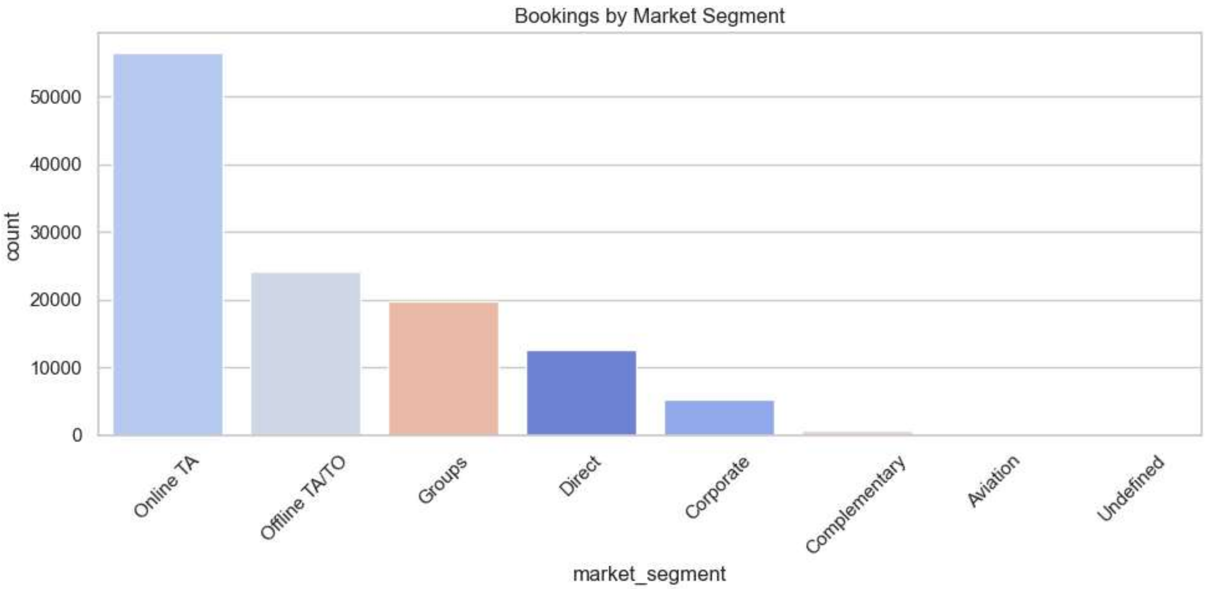
In [23]:

```
# Section 7: Cancellation Rate by Hotel Type ---
cancel_rate = df.groupby('hotel')['is_canceled'].mean()
cancel_rate.plot(kind='bar', color='tomato', figsize=(8, 5))
plt.title('Cancellation Rate by Hotel Type')
plt.ylabel('Cancellation Rate')
plt.xlabel('Hotel')
plt.tight_layout()
plt.show()
```



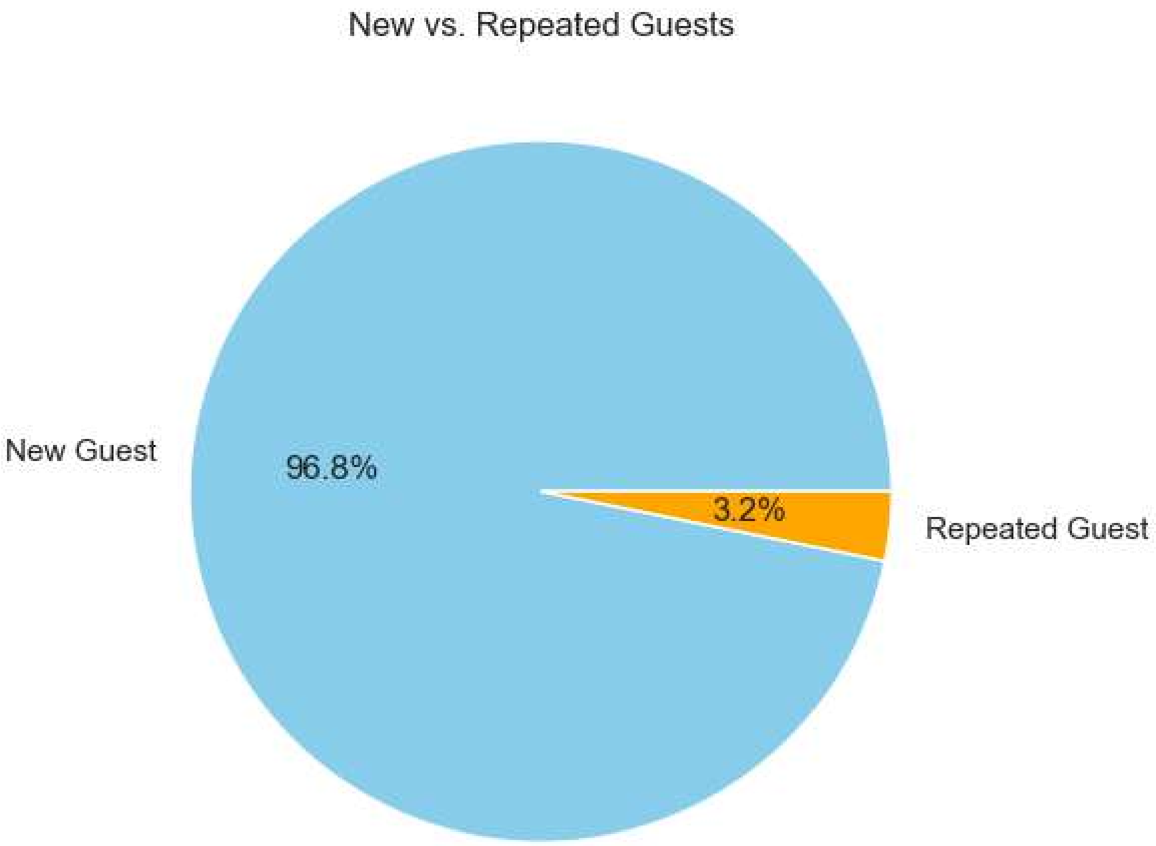
In [21]:

```
# Section 8: Market Segment Analysis ---
plt.figure(figsize=(10, 5))
sns.countplot(x='market_segment', data=df, hue='market_segment', order=df['market_seg
plt.title('Bookings by Market Segment')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



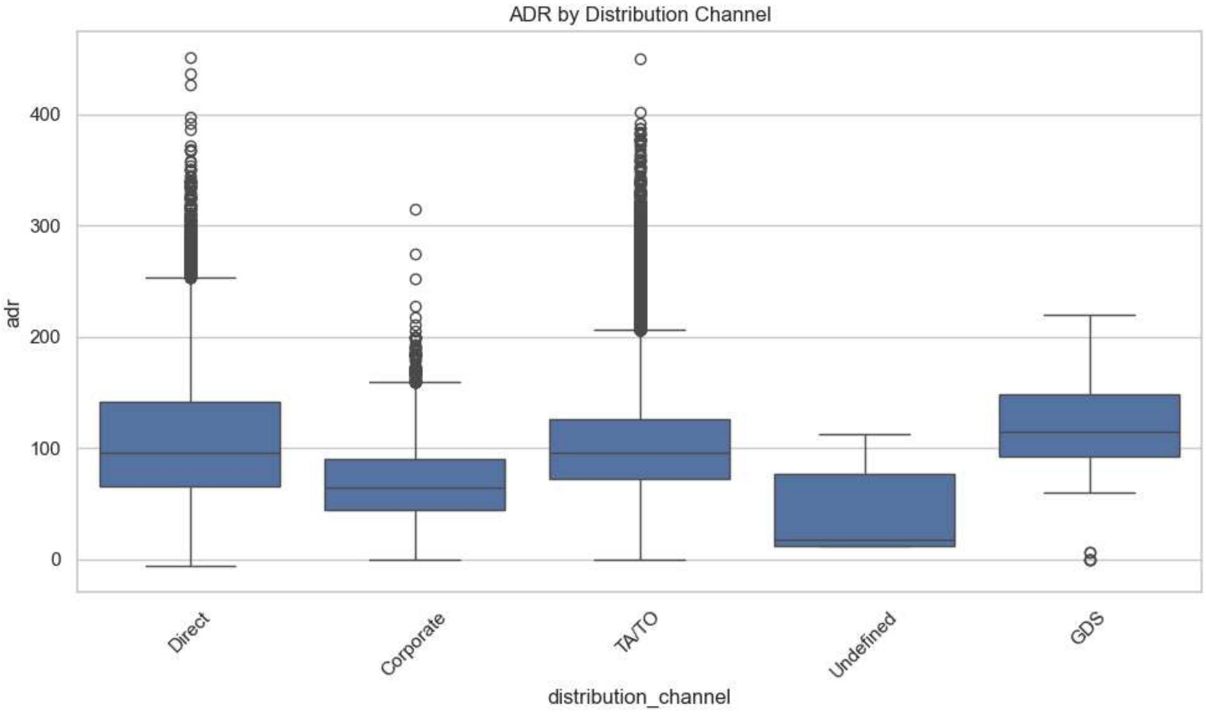
In [22]:

```
#  Section 9: Guest Type Analysis ---
df['guest_type'] = df['is_repeated_guest'].apply(lambda x: 'Repeated Guest' if x == 1
df['guest_type'].value_counts().plot.pie(autopct='%1.1f%%', figsize=(6, 6), colors=['
plt.title('New vs. Repeated Guests')
plt.ylabel('')
plt.tight_layout()
plt.show()
```



In [20]:

```
# Section 10: ADR by Distribution Channel ---
plt.figure(figsize=(10, 6))
sns.boxplot(x='distribution_channel', y='adr', data=filtered_df)
plt.title('ADR by Distribution Channel')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```


ADR by Distribution Channel
adr
400
300
200
100
0
Direct
Corporate
TA/TO
Undefined
GDS
distribution_channel

```
In [ ]:
```