In [20]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import LabelEncoder
```

In [24]:
```
# Load the dataset
file_path = "C:/Users/gundr/Downloads/Day_18_Tours_and_Travels.csv"
df = pd.read_csv(file_path)
```

In [26]:
```
# 1. Handle Missing Values
# Identify missing values
missing_values = df.isnull().sum()
print("Missing Values:\n", missing_values)
```

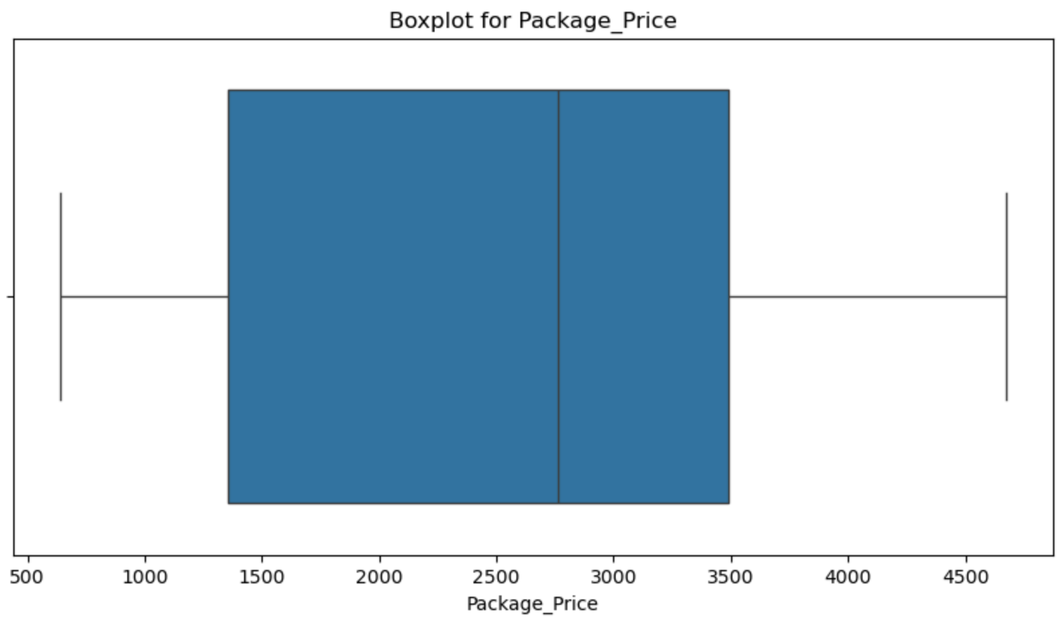
```
Missing Values:
 Booking_ID        0
Destination      27
Package_Price    20
Customer_Age     15
Rating           25
Review_Text      20
Travel_Date      10
dtype: int64
```

In [48]:
```
num_imputer = SimpleImputer(strategy='median')
df[['Customer_Age', 'Rating', 'Package_Price']] = num_imputer.fit_transform(df[['Customer_Age', 'Rating', 'Package_Price']])
```

In [34]:
```
# 2. Detect and Remove Duplicates
df.drop_duplicates(subset=['Review_Text'], keep='first', inplace=True)
```

In [38]:
```
# 4. Identify and Handle Outliers
plt.figure(figsize=(10, 5))
sns.boxplot(x=df['Package_Price'])
plt.title("Boxplot for Package_Price")
plt.show()

plt.figure(figsize=(10, 5))
sns.boxplot(x=df['Rating'])
plt.title("Boxplot for Rating")
plt.show()
```



In [42]:
```
# Removing outliers using IQR for Package_Price
Q1 = df['Package_Price'].quantile(0.25)
Q3 = df['Package_Price'].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR
df = df[(df['Package_Price'] >= lower_bound) & (df['Package_Price'] <= upper_bound)]
```

In [44]:
```
label_encoder = LabelEncoder()
df['Destination'] = label_encoder.fit_transform(df['Destination'])

df['Travel_Date'] = pd.to_datetime(df['Travel_Date'])

df.to_csv("Cleaned_Travel_Reviews.csv", index=False)
```

```
print("Data cleaning completed. Cleaned dataset saved as 'Cleaned_Travel_Reviews.csv'")
```

```
Data cleaning completed. Cleaned dataset saved as 'Cleaned_Travel_Reviews.csv'
```

In [ ]: