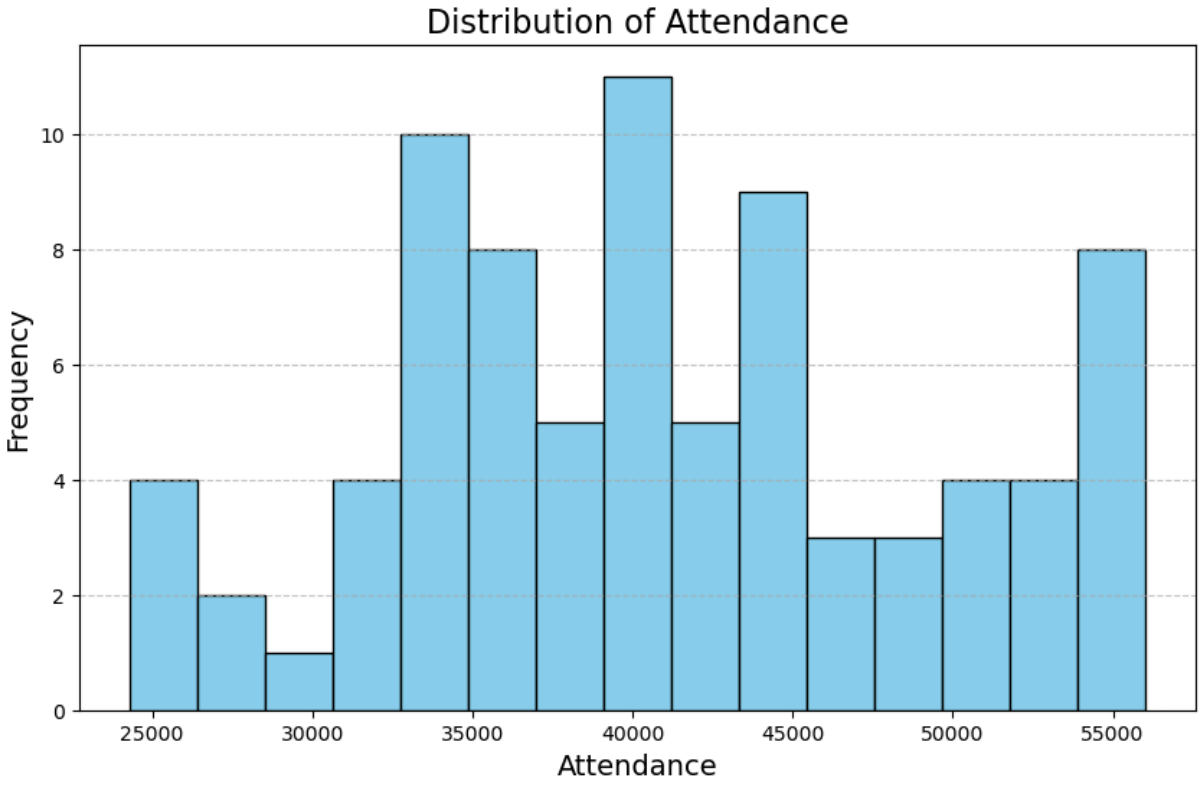
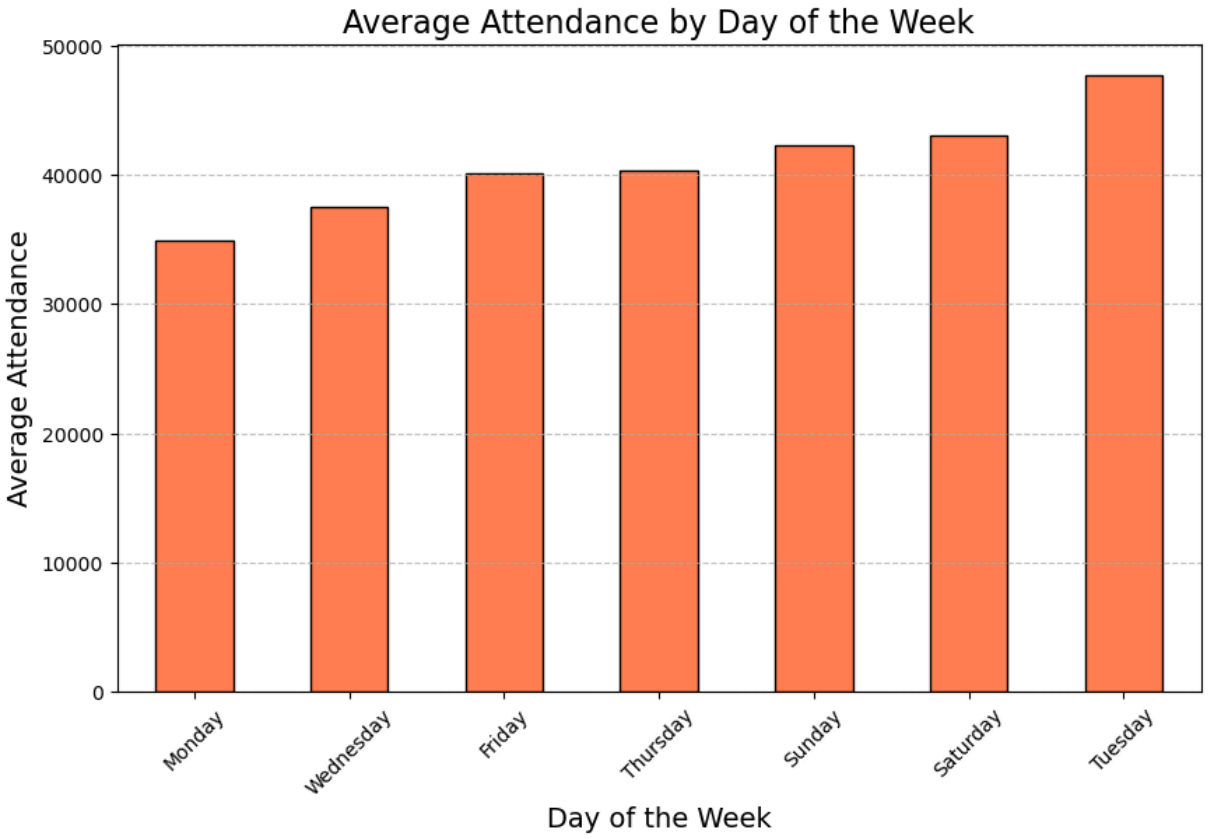
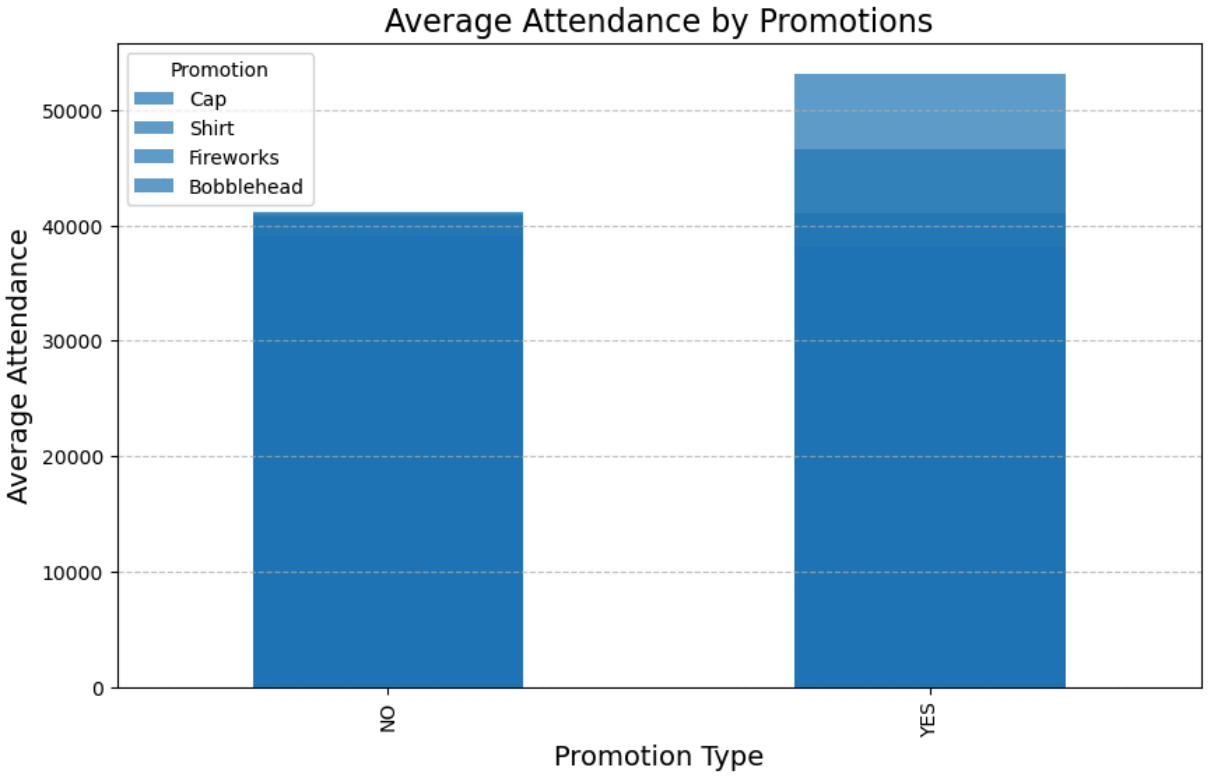
In [ ]:
```
import pandas as pd

# Load dodgers data CSV file to inspect its structure and contents
file_path = r"C:\Users\bobi\Documents\DSC630\hw3\dodgers-2022 (1).csv"
dodgers_data = pd.read_csv(file_path)

# Display the first few rows of the dataset for inspection
dodgers_data.head(), dodgers_data.info()

"""
The dataset has 81 rows and 12 columns, with the following variables:

month: Month of the game.
day: Day of the month.
attend: Attendance at the game.
day_of_week: Day of the week.
opponent: Opponent team.
temp: Temperature during the game.
skies: Weather conditions (e.g., clear, cloudy).
day_night: Whether the game was during the day or at night.
cap: Whether a cap promotion was offered (Yes/No).
shirt: Whether a shirt promotion was offered (Yes/No).
fireworks: Whether there were fireworks (Yes/No).
bobblehead: Whether a bobblehead promotion was offered (Yes/No).


This dataset is clean, with no missing values, and includes relevant factors
"""
```

In [24]:
```
"""
EDA: Explore relationships between attendance and variables such as:
Promotions (cap, shirt, fireworks, bobblehead).
Day of the week.
Opponent team.
Weather and game timing.
Visualization: Create plots to identify patterns in attendance.
Insights and Recommendations: Use findings to suggest strategies to improve
"""


import matplotlib.pyplot as plt

# Analyze attendance distribution
plt.figure(figsize=(10, 6))
plt.hist(dodgers_data['attend'], bins=15, color='skyblue', edgecolor='black'
plt.title('Distribution of Attendance', fontsize=16)
plt.xlabel('Attendance', fontsize=14)
plt.ylabel('Frequency', fontsize=14)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()

# Grouping attendance by day of the week
attendance_by_day = dodgers_data.groupby('day_of_week')['attend'].mean().sor
```

```
plt.figure(figsize=(10, 6))
attendance_by_day.plot(kind='bar', color='coral', edgecolor='black')
plt.title('Average Attendance by Day of the Week', fontsize=16)
plt.xlabel('Day of the Week', fontsize=14)
plt.ylabel('Average Attendance', fontsize=14)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.xticks(rotation=45)
plt.show()

# Grouping attendance by promotions
promotions = ['cap', 'shirt', 'fireworks', 'bobblehead']
promotion_attendance = {promo: dodgers_data.groupby(promo)['attend'].mean()

plt.figure(figsize=(10, 6))
for promo, data in promotion_attendance.items():
    data.plot(kind='bar', label=promo.capitalize(), alpha=0.7)
plt.title('Average Attendance by Promotions', fontsize=16)
plt.xlabel('Promotion Type', fontsize=14)
plt.ylabel('Average Attendance', fontsize=14)
plt.legend(title='Promotion')
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```






In [26]:
```
"""
 initial insights from the exploratory data analysis:

Attendance Distribution:
The distribution shows variability in attendance, with a noticeable number o
likely influenced by specific factors such as promotions or key games.
```

```
Day of the Week:
Attendance tends to be higher on weekends (Saturday and Sunday) compared to
games during the weekend can positively impact attendance.

Promotions:
Promotions like bobbleheads, fireworks, caps, and shirts show a notable incr
Bobblehead promotions, in particular, seem to draw significantly larger crow
"""
```

Out[26]:
```
'\n initial insights from the exploratory data analysis:\n\nAttendance Dist
ribution:\nThe distribution shows variability in attendance, with a noticea
ble number of games reaching high attendance levels, \nlikely influenced by
specific factors such as promotions or key games.\n\nDay of the Week:\nAtte
ndance tends to be higher on weekends (Saturday and Sunday) compared to wee
kdays. This suggests that scheduling \ngames during the weekend can positiv
ely impact attendance.\n\nPromotions:\nPromotions like bobbleheads, firewor
ks, caps, and shirts show a notable increase in average attendance when off
ered. \nBobblehead promotions, in particular, seem to draw significantly la
rger crowds.\n'
```

"""

Next Steps: Investigate other variables such as the impact of weather, temperature, or the opponent team on attendance. Use this information to craft actionable recommendations for the Dodgers' management to improve attendance.

"""