

Johnston630week3

September 20, 2024

Stadium Attendance Analysis

```
[2]: # Imports required packages

import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

pd.set_option('display.max_columns', 50)
pd.set_option('display.max_colwidth', None)
pd.set_option("display.max_rows", 100)
import warnings
warnings.filterwarnings('ignore')
```

```
[3]: # Displays Dataset

df = pd.read_csv("dodgers-2022.csv")

df.head()
```

```
[3]:  month  day  attend  day_of_week  opponent  temp  skies  day_night  cap  shirt  \
0   APR   10   56000     Tuesday    Pirates    67   Clear         Day   NO    NO
1   APR   11   29729   Wednesday    Pirates    58   Cloudy        Night   NO    NO
2   APR   12   28328   Thursday    Pirates    57   Cloudy        Night   NO    NO
3   APR   13   31601    Friday      Padres    54   Cloudy        Night   NO    NO
4   APR   14   46549   Saturday    Padres    57   Cloudy        Night   NO    NO

   fireworks  bobblehead
0          NO          NO
1          NO          NO
2          NO          NO
3         YES          NO
4          NO          NO
```

```
[4]: # Displays dataset statistics
```

```
df.describe()
```

```
[4]:
```

	day	attend	temp
count	81.000000	81.000000	81.000000
mean	16.135802	41040.074074	73.148148
std	9.605666	8297.539460	8.317318
min	1.000000	24312.000000	54.000000
25%	8.000000	34493.000000	67.000000
50%	15.000000	40284.000000	73.000000
75%	25.000000	46588.000000	79.000000
max	31.000000	56000.000000	95.000000

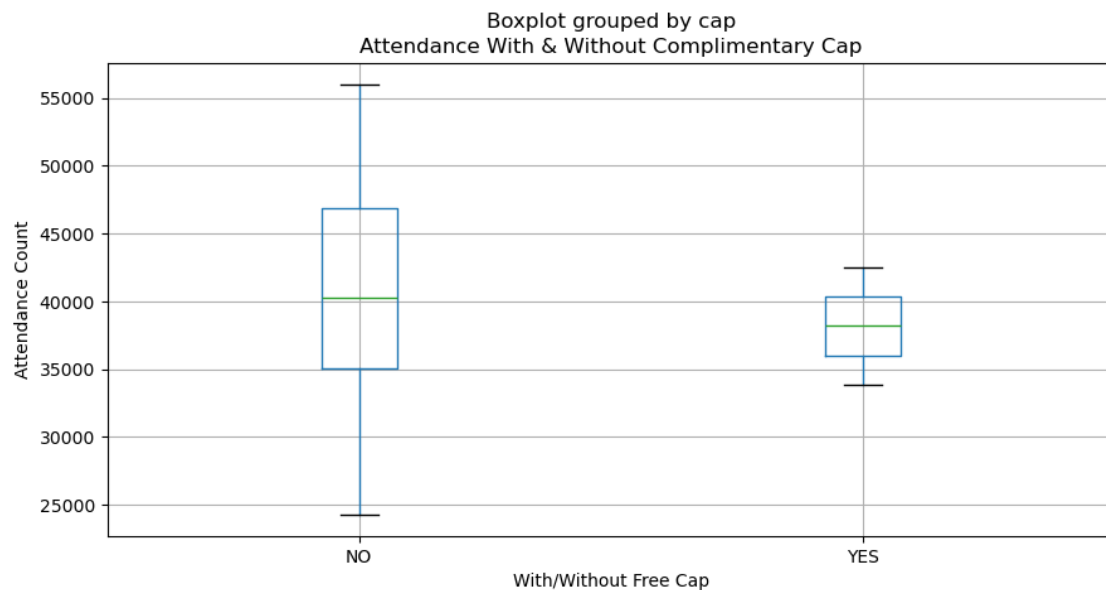
```
[5]: # Scans for null values within the dataset
```

```
df.isnull().values.any()
```

```
[5]: False
```

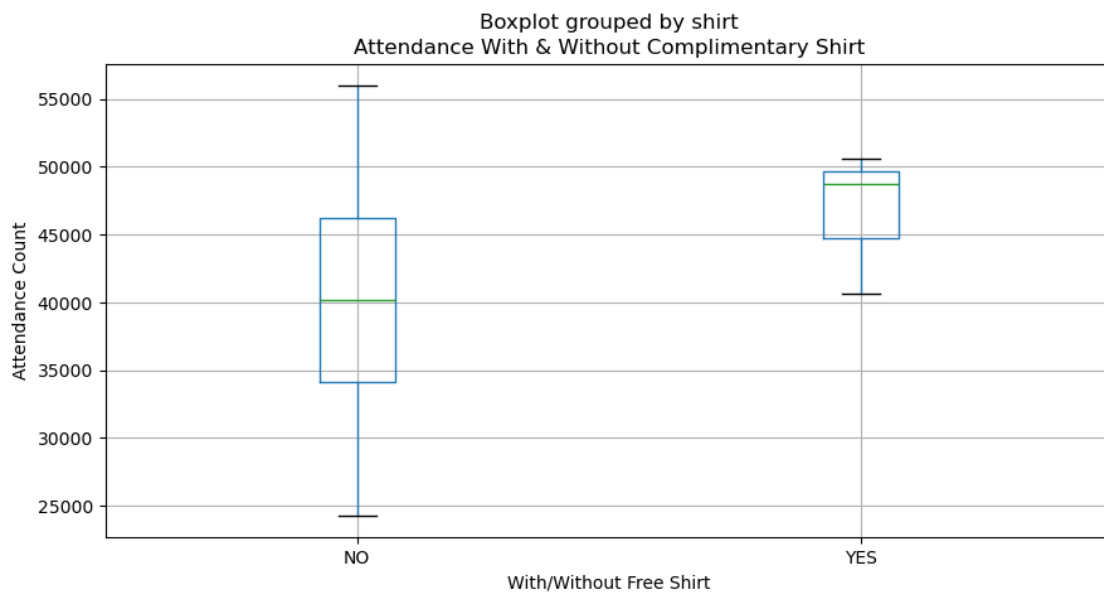
```
[6]: # Boxplot analysis for the cap promotional offer
```

```
df.boxplot(column = 'attend', by = 'cap',figsize = (10,5))  
plt.xlabel("With/Without Free Cap", fontsize = 10)  
plt.ylabel("Attendance Count", fontsize = 10)  
plt.title("Attendance With & Without Complimentary Cap")  
plt.show()
```



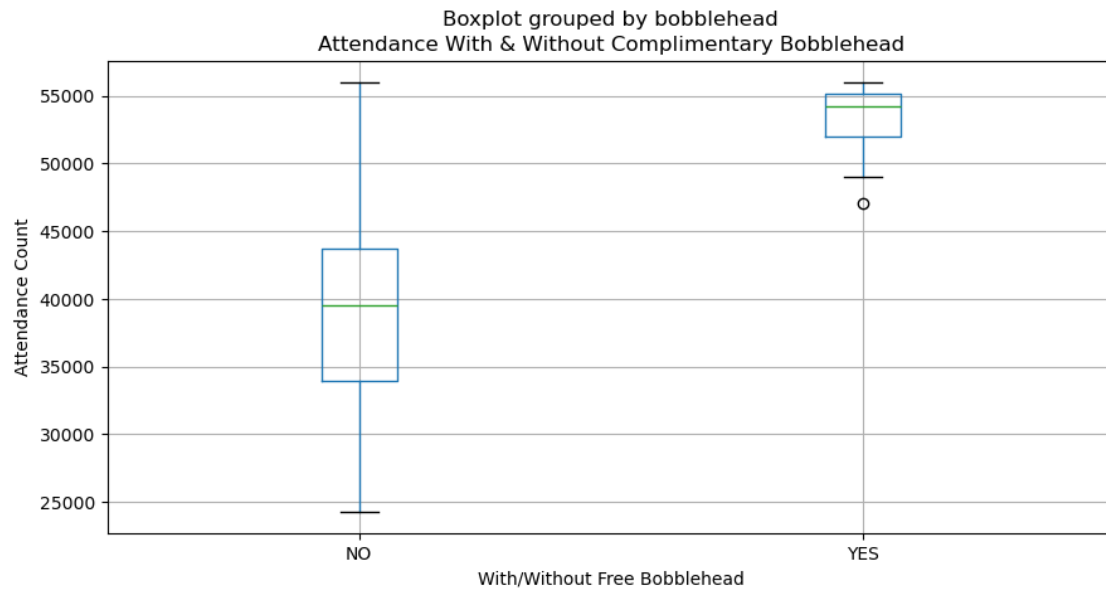
```
[7]: # Boxplot analysis for the shirt promotional offer
```

```
df.boxplot(column = 'attend', by = 'shirt',figsize = (10,5))  
plt.xlabel("With/Without Free Shirt", fontsize = 10)  
plt.ylabel("Attendance Count", fontsize = 10)  
plt.title("Attendance With & Without Complimentary Shirt")  
plt.show()
```



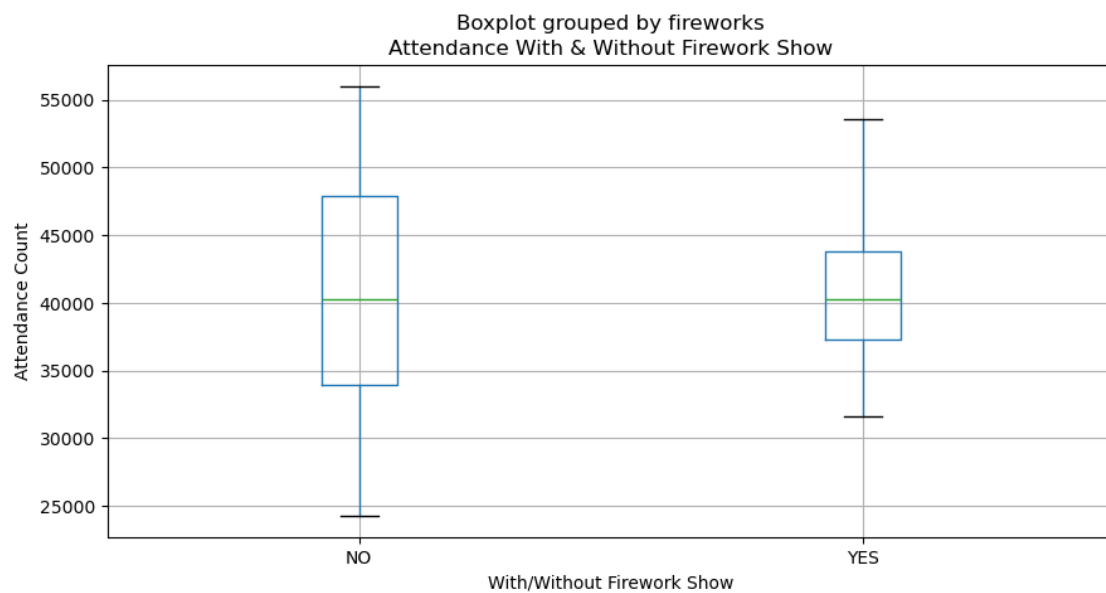
```
[8]: # Boxplot analysis for the bobblehead promotional offer
```

```
df.boxplot(column = 'attend', by = 'bobblehead',figsize = (10,5))  
plt.xlabel("With/Without Free Bobblehead", fontsize = 10)  
plt.ylabel("Attendance Count", fontsize = 10)  
plt.title("Attendance With & Without Complimentary Bobblehead")  
plt.show()
```



[9]: *# Boxplot analysis for the firework show promotional offer*

```
df.boxplot(column = 'attend', by = 'fireworks', figsize = (10,5))
plt.xlabel("With/Without Firework Show", fontsize = 10)
plt.ylabel("Attendance Count", fontsize = 10)
plt.title("Attendance With & Without Firework Show")
plt.show()
```



```
[10]: # creates a subset of target variables
```

```
subset = df[["attend", "temp", "day_of_week", "opponent", "month"]]  
  
subset.head()
```

```
[10]:
```

	attend	temp	day_of_week	opponent	month
0	56000	67	Tuesday	Pirates	APR
1	29729	58	Wednesday	Pirates	APR
2	28328	57	Thursday	Pirates	APR
3	31601	54	Friday	Padres	APR
4	46549	57	Saturday	Padres	APR

```
[11]: # Displays attendance by month
```

```
monthgroup = subset.groupby(['month']).mean(numeric_only = True)['attend'].  
    ↪sort_values().reset_index()  
monthgroup
```

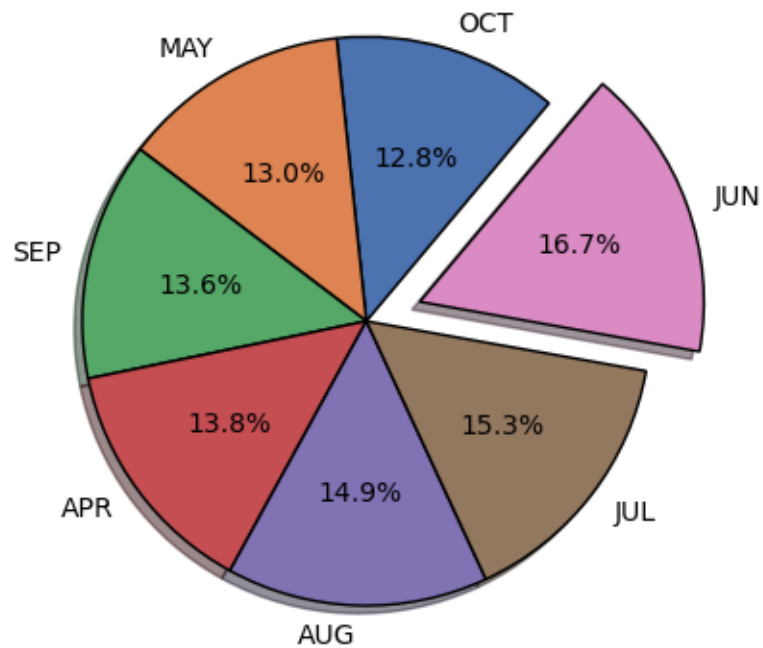
```
[11]:
```

	month	attend
0	OCT	36703.666667
1	MAY	37345.722222
2	SEP	38955.083333
3	APR	39591.916667
4	AUG	42751.533333
5	JUL	43884.250000
6	JUN	47940.444444

```
[12]: # Displays pie chart of attendance by month
```

```
explode = [0, 0, 0, 0, 0, 0, 0.2]  
  
plt.pie(monthgroup["attend"], labels = monthgroup["month"], explode = explode,   
    ↪shadow = True,  
startangle = 50,  
    autopct = "%1.1f%%",  
    colors = sns.color_palette("deep"),  
    wedgeprops = {'edgecolor': 'black'})  
  
plt.title("Yearly Attendance Percentage by Month")  
plt.figure(figsize = (20, 10))  
plt.show()
```

Yearly Attendance Percentage by Month



<Figure size 2000x1000 with 0 Axes>

[13]: *# Displays attendance by each day of the week*

```
daysgroup = subset.groupby(['day_of_week']).mean(numeric_only = True)['attend'].
    ↪sort_values().reset_index()
daysgroup
```

```
[13]:   day_of_week      attend
0    Monday  34965.666667
1  Wednesday  37585.166667
2    Friday  40116.923077
3  Thursday  40407.400000
4    Sunday  42268.846154
5  Saturday  43072.923077
6    Tuesday  47741.230769
```

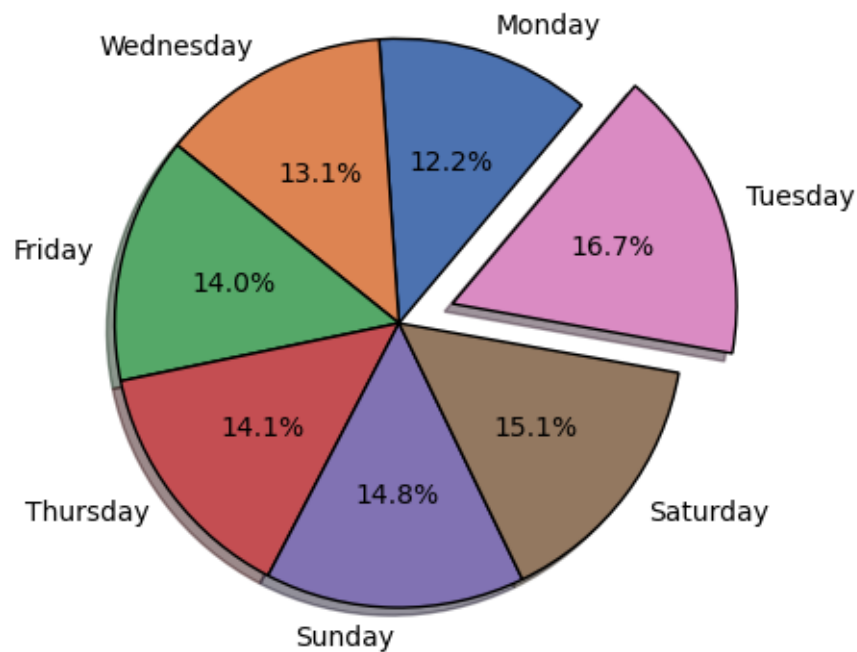
[14]: *# Displays pie chart of attendance by each day of the week*

```
explode = [0, 0, 0, 0, 0, 0, 0.2]
```

```
plt.pie(daysgroup["attend"], labels = daysgroup["day_of_week"], explode =
    explode, shadow = True,
startangle = 50,
    autopct = "%1.1f%%",
    colors = sns.color_palette("deep"),
    wedgeprops = {'edgecolor': 'black'})

plt.title("Yearly Attendance Percentage by Day of the week")
plt.figure(figsize = (20, 10))
plt.show()
```

Yearly Attendance Percentage by Day of the week



<Figure size 2000x1000 with 0 Axes>

```
[15]: # Displays occurances of each attendance and temperature count along with
# their median values

pair1 = ["attend", "temp"]

for inx,col in enumerate(pair1):

    median_line = subset[col].median()
    plt.figure(figsize = (8, 4))
```

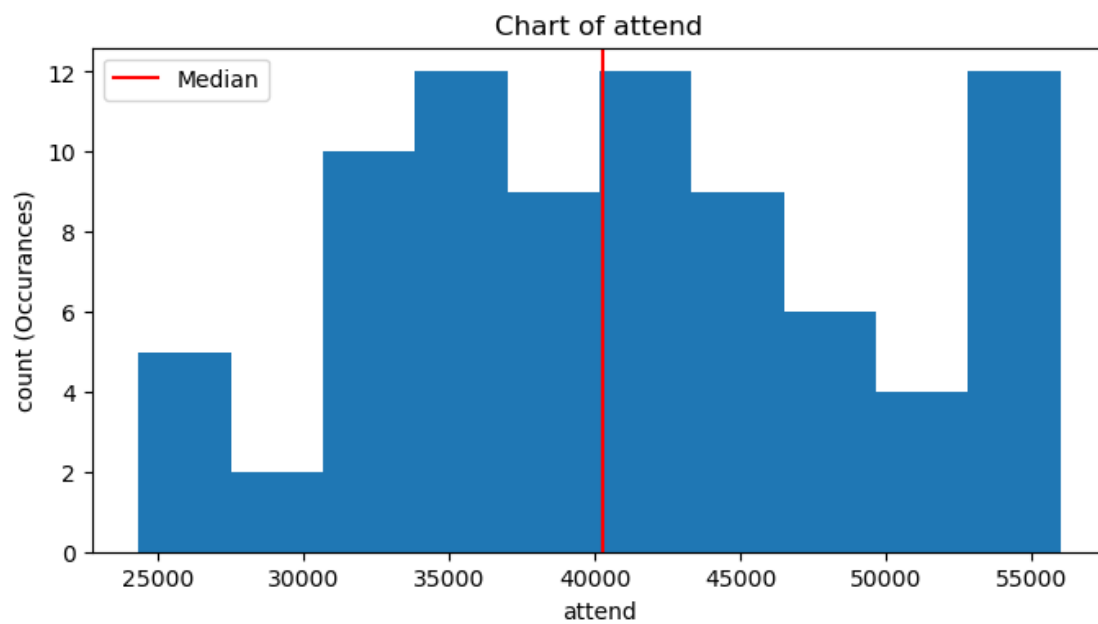
```

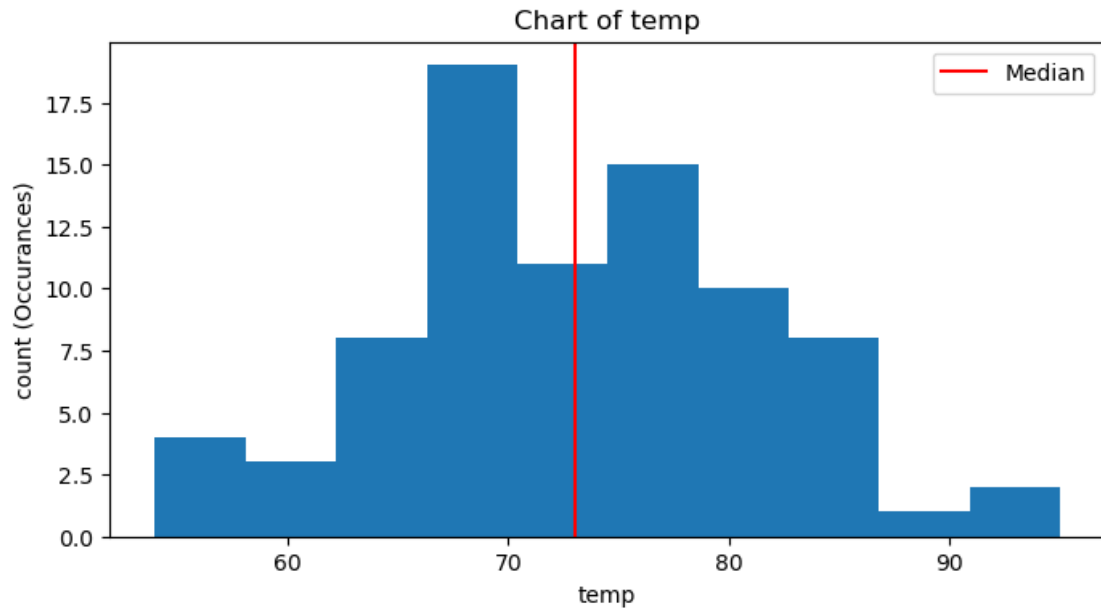
color = 'Red'
print(f"The Median of {col} is {median_line}")
plt.axvline(median_line, color = color, label = 'Median')
subset[col].hist(grid = False)
plt.xlabel(subset[col].name)
plt.ylabel('count (Occurrences)')
plt.legend()
plt.title('Chart of ' + col)
plt.show

```

The Median of attend is 40284.0

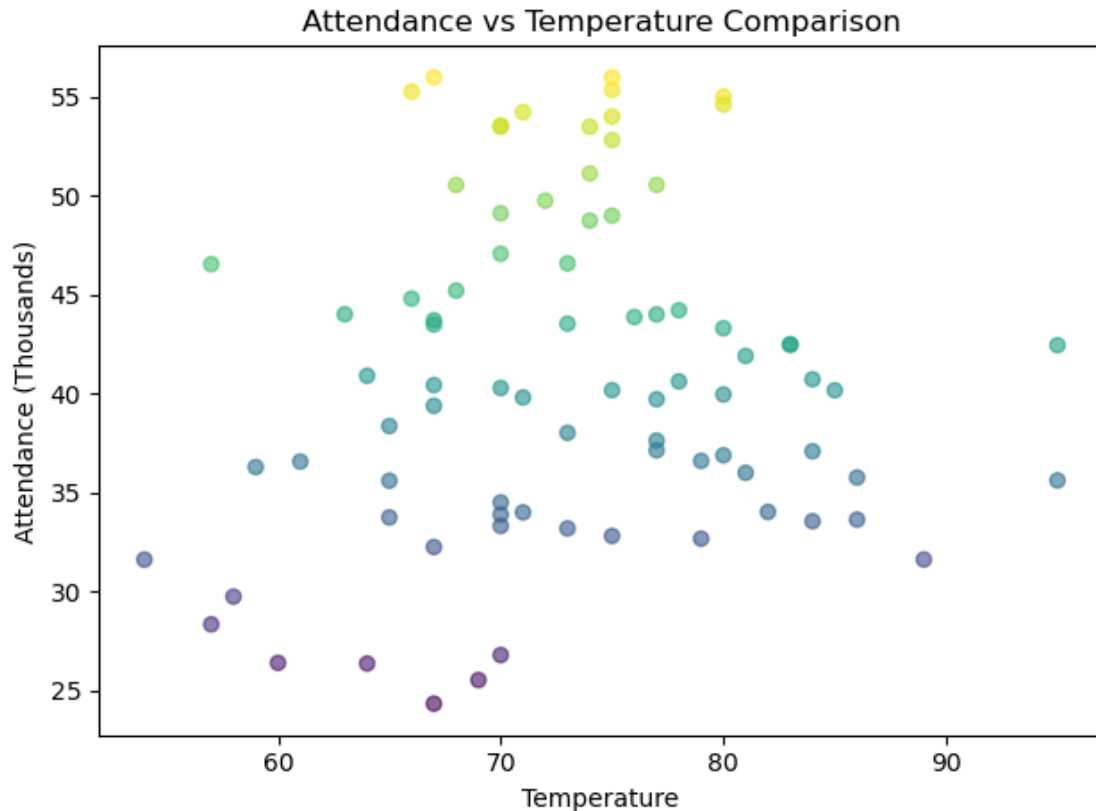
The Median of temp is 73.0





```
[16]: # Scatterplot analysis of Temperature vs Attendance

plt.scatter(subset["temp"],subset["attend"]/1000, s=35, alpha=0.6, c =_
↳subset["attend"]
/1000)
plt.title("Attendance vs Temperature Comparison")
plt.xlabel('Temperature')
plt.ylabel('Attendance (Thousands)')
plt.tight_layout()
plt.show()
```

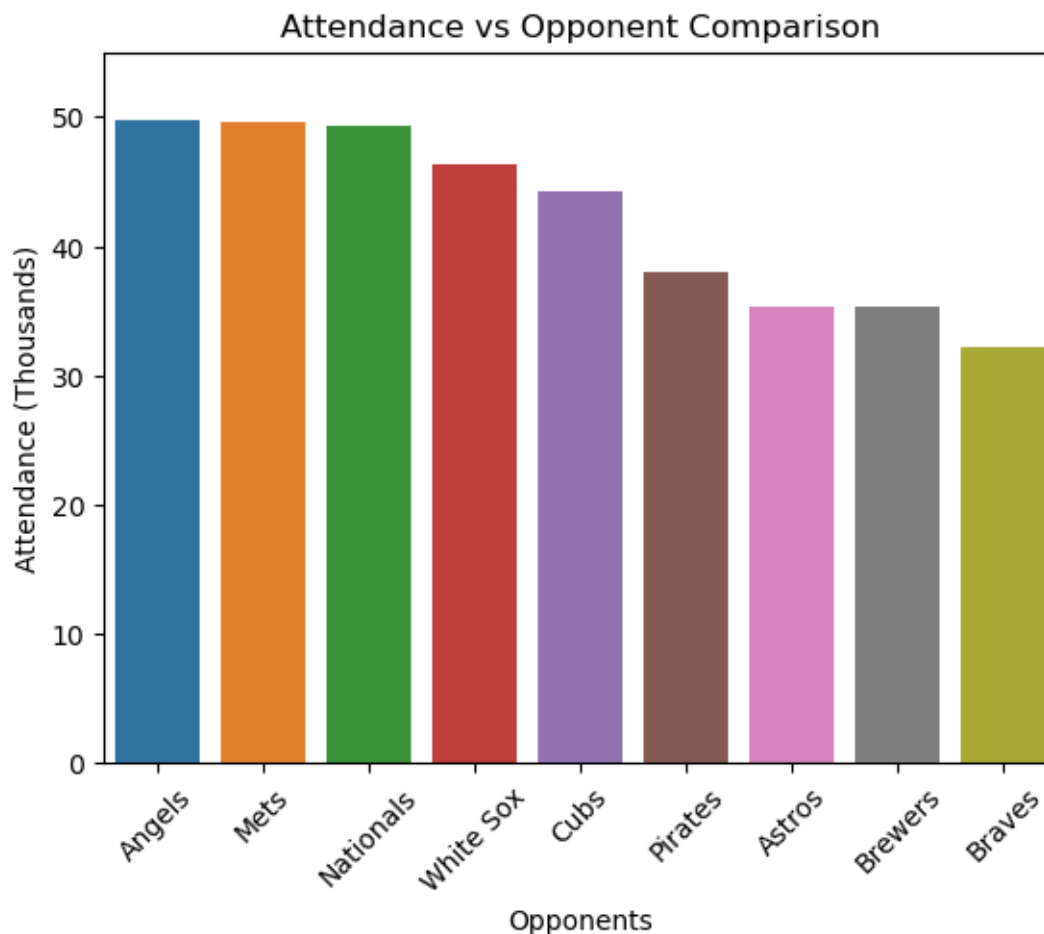


```
[17]: # Sets group of each opponents mean attendance count

opponentsgroup = subset.groupby('opponent').mean(numeric_only = True)['attend'].
    ↪round().reset_index().sort_values(by =
['attend'], ascending = False)
opponentsgroup["attend"] = opponentsgroup["attend"]/1000
opponentset = pd.concat([opponentsgroup[:5], opponentsgroup[-4:]], ignore_index=
    ↪ True, axis = 0)
```

```
[18]: # Displays each opponents mean attendance count

sns.barplot(x = 'opponent',y = 'attend',data = opponentset, estimator = np.mean)
plt.title('Attendance vs Opponent Comparison')
plt.xlabel('Opponents')
plt.ylabel('Attendance (Thousands)')
plt.xticks(rotation = 45)
plt.ylim(0, 55)
plt.show()
```



Final Report/Review

Assumptions

My analysis will not include the following columns : day_night, skies and day. My reasons for this is that I do not believe that anyone will choose not to go to a game because its being played on Saturday - Nov 6th but will choose to go to a game if its on Saturday - Nov 11th. The forecast is something that at best can be predicted a few days in advance and sports schedules are finalized before the season begins so I believe skies is a somewhat redundant statistic. Day games will have a lower attendance during the week than night games due to work/school schedules and this will be the opposite during the weekend. Also, the time of day that a game is being played is decided by the TV provider that has the deal with the league in which the sport is played. For example, FOX/CBS with the NFL and Sky Sports with the Premier League.

Analysis

The areas of concern in the attendance record are temperature outside of 68-82 degrees, Monday/Wednesday games, May/October games and games against opponents that aren't either local/historical rivals or holding good records i.e(Angels/Astros). Promotional offers are one way to control the fight against low attendances. Because the dodgers stadium does not have a retractable

roof temperature control is difficult unless they wish to add headed seats as the majority of negative outliers are in temperatures below 70 degrees. To fit this idea to the entire stadium would be likely cost ineffective so I will recommend fighting the previously mentioned areas of importance with promotional offers or negotiations with the TV companies who control match start times/scheduling. The promotional offers with positive average outcomes are bobbleheads and shirts. Caps and Fireworks show little to no improvement towards the attendance records. Bobbleheads in particular have an extremely high average influence on game attendance. I would advise that throughout the season all game tickets are presented with the choice of a complimentary shirt or bobblehead on Mondays and Wednesdays. I would also offer a similar offer throughout the months of May and October. I understand that it is a project that would cost MILLIONS but as I stated earlier as the majority of negative attendance occurrences in regards to temperature are in sub 70 degree weather a retractable roof for the stadium would be a great success. This is also a feature that a handful of teams also already have, it would be smart to stay up-to-date with the competition. A roof would allow for an added element of temperature control and would shelter consumers from cold winds. In an ideal world I would suggest to request that games against certain opponents be played in the months in question but I understand that the TV companies/MLB board of commissions would not make any special exceptions for a specific team. So again I would complimentary merchandise with the purchase of any team not based locally or with a sub .500 record at the time of play. I would leave the firework shows for games that are predicted to have a high level of attendance since they have the lowest level of influence on attendance. I would use the Cap promotion in games played where the temperature is $90 <$ degrees as that statistic also shows negative implications and would offer the consumers an element of shade/cover from the sun. The cap promotion actual has a slightly negative correlation to game attendance so I would use it sparingly. Although, this statistic is likely an outlier for this specific season as I can't imagine someone deciding to not attend a game because they received a free hat.

This concludes my strategy for implementing the promotions we have on offer and recommended possible future alterations to the Dodgers stadium. The average attendance for this season was 41,040/56000 seats or a 73% stadium capacity. With these recommendations I believe that it would be reasonable to predict an increase to 45,000/56000 which would be a 7% increase in stadium capacity to 80% total. The average price of a Dodgers ticket is \$134.73, this increase would see on average an extra 3,960 game day attendees. This would result in game day revenue increasing by \$533,530.80. This is also ONLY based upon ticket sales and does not take into consideration the expected increase in concessions or merchandise.

Reference(s):

- Buy Dodgers tickets - Los Angeles Dodgers MLB tickets at Ticketsmarter. ticketsmarter.com. (n.d.). <https://www.ticketsmarter.com/p/los-angeles-dodgers-tickets#:~:text=For%20the%20current%20season%2C%20Los,2024%20season%20is%20around%20%24134.7>