3.2 Exercise.DSC630 - Jennifer Barrera Conde

June 19, 2024

1 Exercise 3.2

2 DSC630

3 Jennifer Barrera Conde

4 Recommendation for Attendance Improvement

To analyze and make recommendations to improve attendance for the Los Angeles Dodgers Major League Baseball (MLB), we need to explore the given data to identify factors that influence game attendance.

NOTE: My plan of action is to first get to know the data and present key findings at the end. There may not be much information on each cell where I explain what I see because I aim to work it as if giving a presentation to my management team with the aid of the visuals I'll be creating. I am doing this to avoid redundance and repetitiveness.

[ ]:

4.1 Step 1. Getting to know the data and cleaning:

Load the data and check for any missing or inconsistent values. Ensure all data types are appropriate for analysis.

[1]:
```
import pandas as pd

# Load the data
data = pd.read_csv('dodgers-2022.csv')

# Display the first few rows of the data
data.head()
```

[1]:

	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
```

[2]:
```
# Check for missing values and data types
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 81 entries, 0 to 80
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype 
---  ------       --------------  ----- 
 0   month        81 non-null     object
 1   day          81 non-null     int64 
 2   attend       81 non-null     int64 
 3   day_of_week  81 non-null     object
 4   opponent     81 non-null     object
 5   temp         81 non-null     int64 
 6   skies        81 non-null     object
 7   day_night    81 non-null     object
 8   cap          81 non-null     object
 9   shirt        81 non-null     object
 10  fireworks    81 non-null     object
 11  bobblehead   81 non-null     object
dtypes: int64(3), object(9)
memory usage: 7.7+ KB
```

The dataset includes various features such as the date, day of the week, opponent, weather conditions, promotional events, and attendance numbers.

There are different ways I could work the data such as: 1. Generate summary statistics for attendance. 2. Relationship between attendance and day of the week. 3. Impact of promotional events (example fireworks, bobblehead giveaways). 4. Influence of opponent teams and weather conditions on attendance. 5. Determine the impact of the weather conditions on attendance (using temperature and skies).

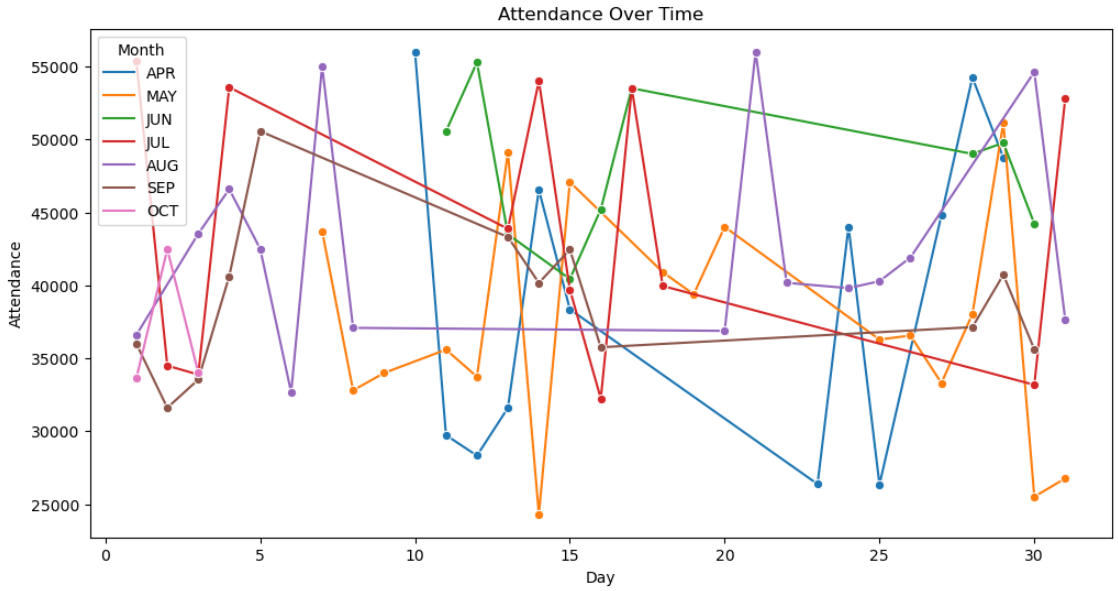
[ ]:

4.2 Step 2. Descriptive Statistics and Visualization:

Generate summary statistics for the attendance and other numerical features. Visualize the data to identify trends and patterns, such as attendance over time.

[3]:

```
#  Compute descriptive statistics for the attendance and visualize some initial␣
 ↪patterns
import matplotlib.pyplot as plt
import seaborn as sns

# Summary statistics for attendance
attendance_stats = data['attend'].describe()
attendance_stats

# Plot attendance over time
plt.figure(figsize=(12, 6))
sns.lineplot(data=data, x='day', y='attend', hue='month', marker='o')
plt.title('Attendance Over Time')
plt.xlabel('Day')
plt.ylabel('Attendance')
plt.legend(title='Month')
plt.show()
```

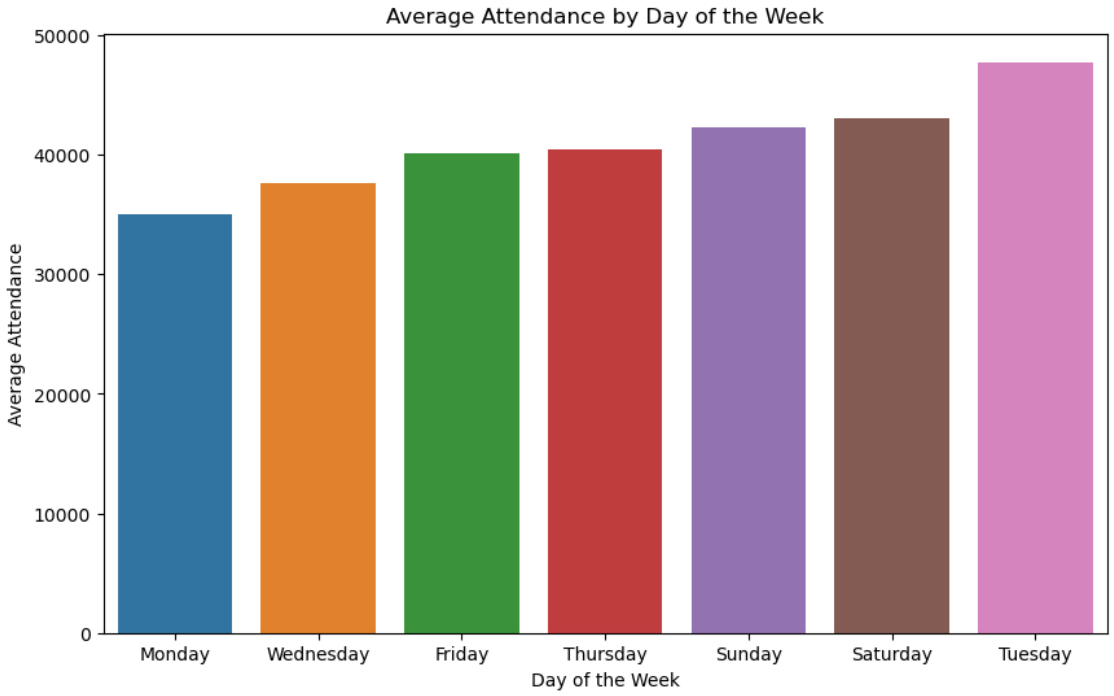


[ ]:

4.3 Step 3. Perform an Exploratory Data Analysis (EDA):

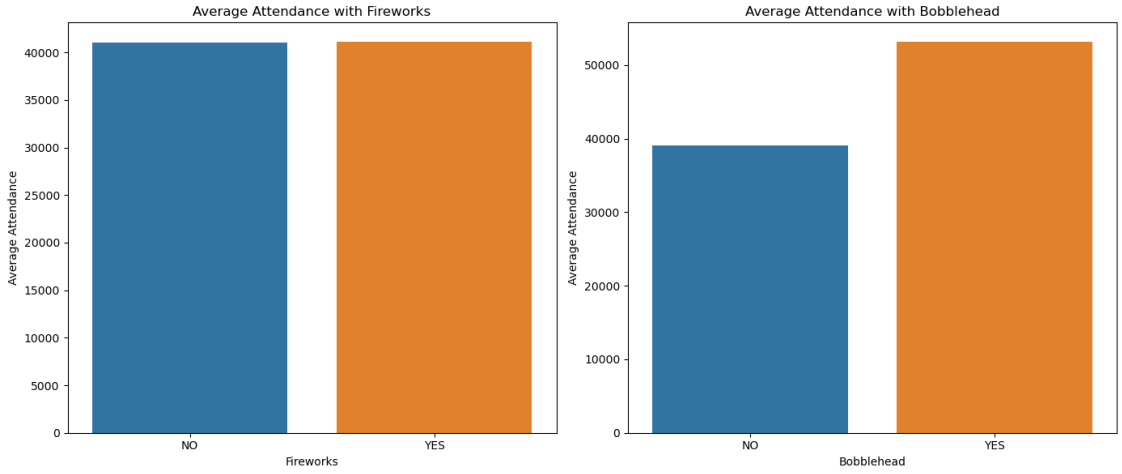
Investigate the relationship between attendance and day of the week. Analyze the impact of promotional events (example, fireworks, bobblehead giveaways). Examine the influence of opponent teams and weather conditions on attendance.

[4]:
```
# Average attendance by day of the week
avg_attendance_by_day = data.groupby('day_of_week')['attend'].mean().
 ↪sort_values()

plt.figure(figsize=(10, 6))
sns.barplot(x=avg_attendance_by_day.index, y=avg_attendance_by_day.values)
plt.title('Average Attendance by Day of the Week')
plt.xlabel('Day of the Week')
plt.ylabel('Average Attendance')
plt.show()
```



[5]:
```
# Analyze the impact of promotional events like fireworks and bobblehead␣
 ↪giveaways.
# Average attendance with and without fireworks
avg_attendance_fireworks = data.groupby('fireworks')['attend'].mean()

# Average attendance with and without bobblehead
avg_attendance_bobblehead = data.groupby('bobblehead')['attend'].mean()

fig, axes = plt.subplots(1, 2, figsize=(14, 6))

sns.barplot(x=avg_attendance_fireworks.index, y=avg_attendance_fireworks.
 ↪values, ax=axes[0])
axes[0].set_title('Average Attendance with Fireworks')
```

```
axes[0].set_xlabel('Fireworks')
axes[0].set_ylabel('Average Attendance')

sns.barplot(x=avg_attendance_bobblehead.index, y=avg_attendance_bobblehead.
 ↪values, ax=axes[1])
axes[1].set_title('Average Attendance with Bobblehead')
axes[1].set_xlabel('Bobblehead')
axes[1].set_ylabel('Average Attendance')

plt.tight_layout()
plt.show()
```



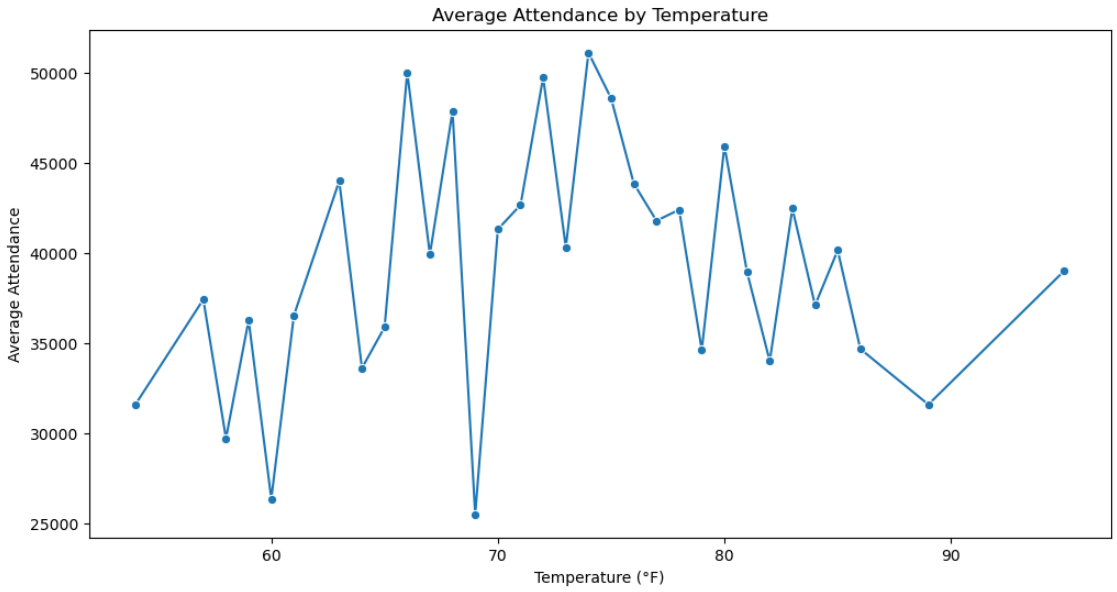
[ ]:

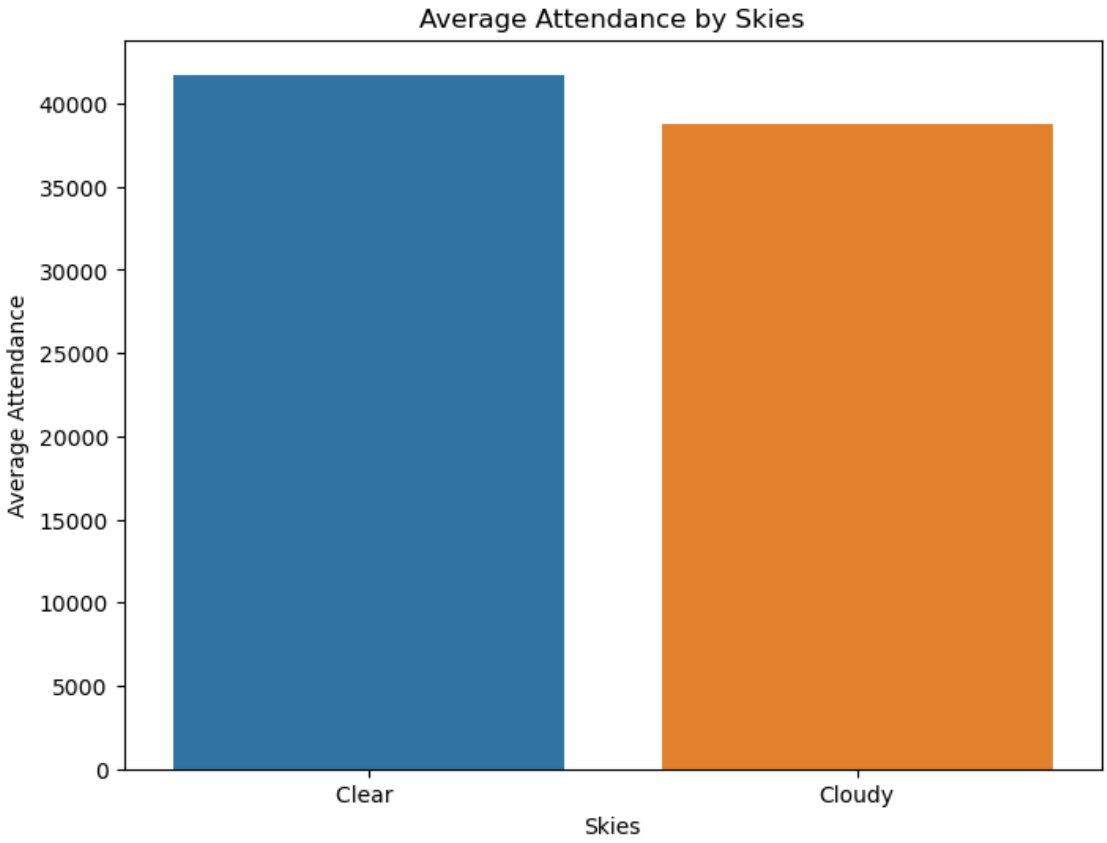
4.4 Step 4. Modeling and Hypothesis Testing:

Use regression analysis to quantify the impact of various factors on attendance. Test hypotheses to validate which factors significantly influence attendance.

[6]:
```
# Determine the impact of the weather conditions on attendance
# Average attendance by temperature
avg_attendance_temp = data.groupby('temp')['attend'].mean()

plt.figure(figsize=(12, 6))
sns.lineplot(x=avg_attendance_temp.index, y=avg_attendance_temp.values,␣
 ↪marker='o')
plt.title('Average Attendance by Temperature')
plt.xlabel('Temperature (°F)')
plt.ylabel('Average Attendance')
plt.show()
```

```
# Attendance based on skies (Clear vs Cloudy)
avg_attendance_skies = data.groupby('skies')['attend'].mean()

plt.figure(figsize=(8, 6))
sns.barplot(x=avg_attendance_skies.index, y=avg_attendance_skies.values)
plt.title('Average Attendance by Skies')
plt.xlabel('Skies')
plt.ylabel('Average Attendance')
plt.show()
```





[ ]:

4.5 Step 5. Recommendations:

Based on the findings, provide actionable recommendations to management. Let's start by loading and examining the data.

4.6 Key Findings and Recommendations:

4.6.1 Day of the Week:

Attendance tends to be higher on Tuesdays, followed by the weekends as second highest. Management should schedule more high-profile games or promotional events on these days to maximize attendance. There could also be the option to attract attendees if some mid to high profile games were to take place during the week. This with the goal that attendance could be spread and made into several days of the week rather than only 3 to 4 strong days of the week.

4.6.2 Promotional Events:

Games with fireworks and bobblehead giveaways significantly boost attendance. Increasing the frequency of such events could draw larger crowds. For example, consider adding more bobblehead

giveaway nights and aligning them with popular to mid opponents. I would make a second bobblehead day on the lowest attended games which are Mondays, or Thursdays since it spreads out the crowd more, from having back to back promotional events. Fridays have good attendance, but attendees are indiferent of the firworks, whilst Tuesday attendance is the highest with a significant reasoning being the bobbleheads.

4.6.3 Opponent Teams:

Games against high-profile opponents (e.g., Giants, Cubs) generally see higher attendance. Scheduling more games with popular rivals and promoting these matchups can increase interest.

4.6.4 Weather Conditions:

Clear skies correlate with higher attendance. While weather cannot be controlled, management can use this insight for better forecasting and planning.

4.6.5 Temperature:

Moderate temperatures (70-80°F) are associated with higher attendance. Early and late-season games might need additional promotions to counteract less favorable weather conditions due to the location.

4.6.6 Conclusion:

To improve attendance, the Dodgers management should focus on scheduling more weekend games, increasing promotional events (especially bobblehead nights), and leveraging high-profile opponents. These strategies, supported by the data analysis, will likely lead to higher game attendance and enhanced fan engagement. There is also the option of coming up with other promotional events such as limited release or limited edition items on the lowest attended games. In my opinion, attendees are attracted to physical items due to baseball being a very highly collectible sport. Many baseball fans have collected baseballs, bats, bobbleheads, cards, and such due to Baseball being a strong part of American history with beloved fans all over the world that can see their collectible and have nostalgia from that special time they spent with friends, family, coworkers, and more.

Story that can be skipped: I was taken to my first baseball game last year (2023), I have never experienced baseball this way before even though it was just a Double-A Affiliate of the Miami Marlins I had the time of my life, so much so that I have 3 pieces of memorabilia for keepsake. I have a baseball ball of the team which I tried so hard to catch one, but couldn't and someone gifted me one. I have a cap of the MiLB Pok-Ta-Pok diversion cup, which I am Mexican so I **had** to have it. I also have stickers and a postcard/calendar. There were many people purchasing items and collecting them to not just make memories but to save them and make those items a recolection of what they been through. Which is why I believe that those promotional games where one can get a physical item for their keepsake will definitely attract more attendees. I can say so from experience. I keep wanting to go to more games and more teams and experience it all over again.

[ ]: