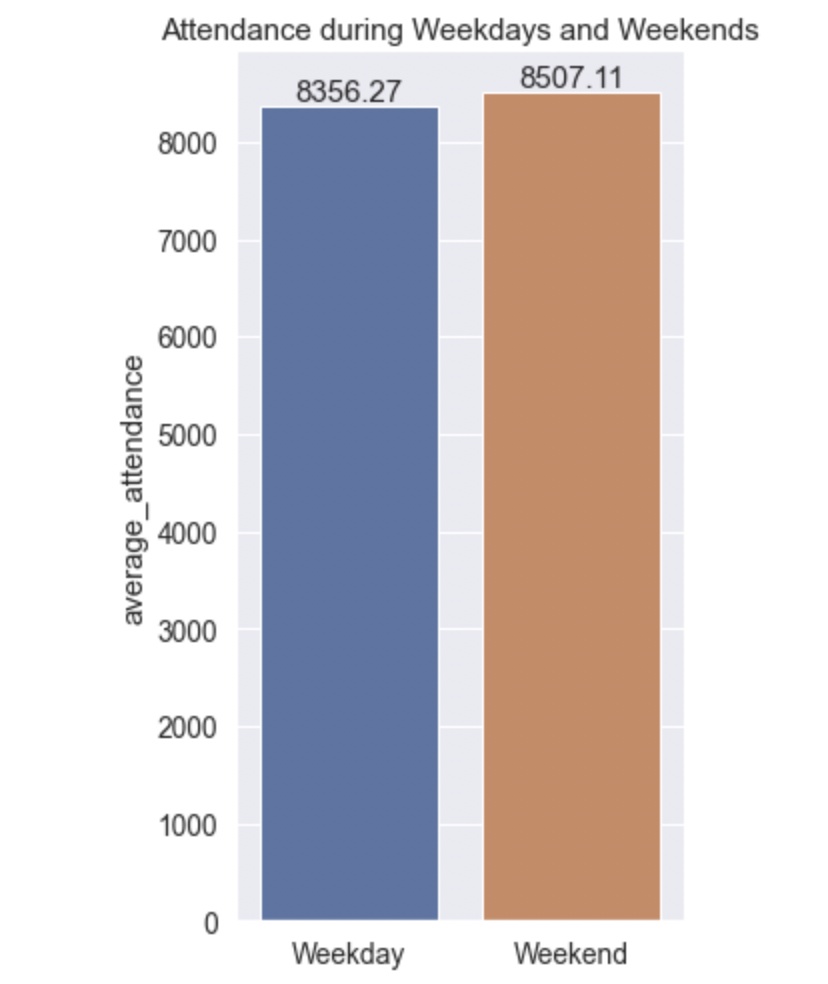


Q1) **[SQL]** Do weekend games (Friday, Saturday, Sunday), on average, have a higher attendance than weekday games (Monday through Thursday)?

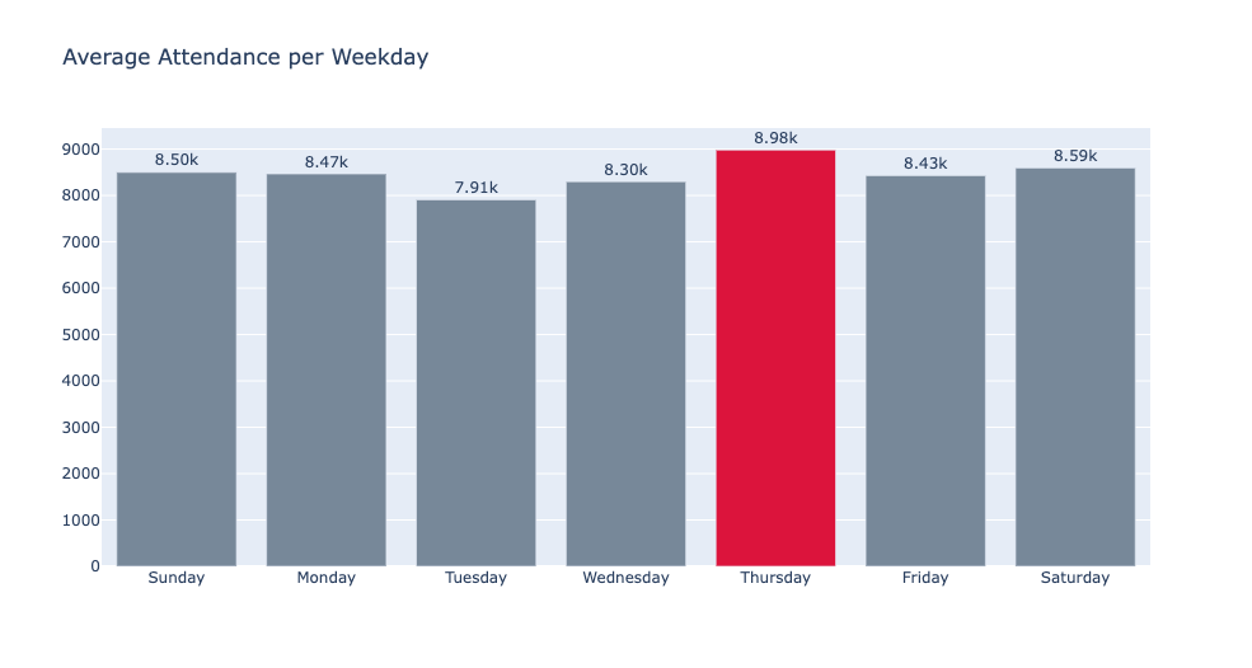
**Methodology :**

1. Check for Null or duplicate values on SCAN dataset
2. Converted the scan\_datetime column to Pacific Time and extracted the day of the week from this date.
3. Found total scans per event using the COUNT function and the groupby function.
4. Calculated the Mean of total scans for each day of the week and graphed a barplot.
5. Calculated the Mean of total scans for weekdays and weekends.
6. Visualized the total attendance using a Barplot.

**Insight :** Looking at the graph, it plots the average attendance on weekdays and weekends. As expected, we see that the average attendance on Weekends is around 8500 and on Weekdays attracts fewer(~200 lesser) crowds.



**Insight :** Having a look at the average attendance throughout the week, Thursdays are surprisingly the most crowded. While Tuesdays have least attendance, it is seen that the attendance on the rest of the days is almost around 8500.

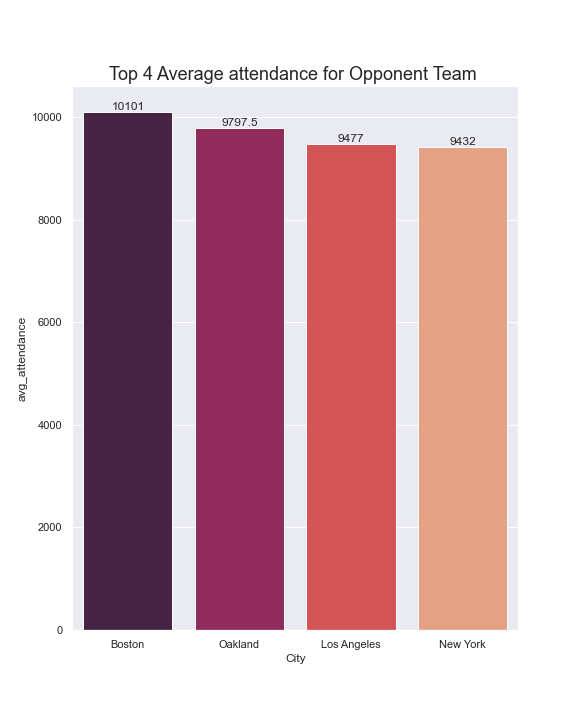


Q2) **[SQL]** Identify and rank the top 4 opponents with the highest average number of attendances.

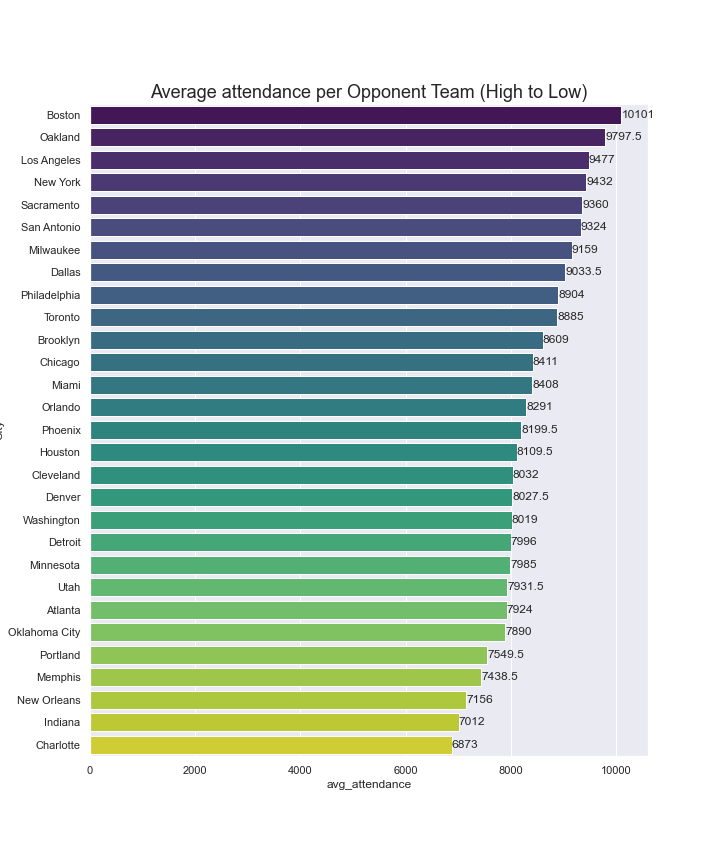
**Methodology :**

1. Using the SCAN dataset, calculate the total attendance by grouping the event\_name column..
2. Extract City name from event\_name column.
3. Find the mean of total attendance of each city.
4. Plot the total attendance for each city to find the top 4 highest average number of attendances for every opponent.

**Insight :** Having a look at the highest average attendance for the top 4 teams in descending order, Boston tops the list. Golden State is next followed by Los Angeles and then New York. By insight it seems that people want to watch the “popular” teams that have more Championships.



**Insight :** Here, I have plotted the teams with average attendance for opponent teams in descending order. It varies from around 7000 for Charlotte to 10000 for Boston.

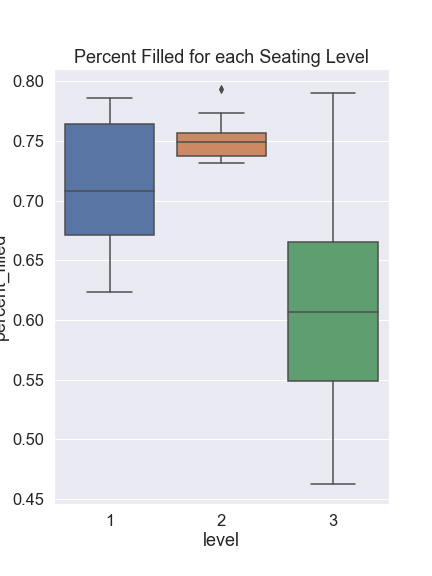


Q3) Identify and rank the top 10 sections that are, on average, the most filled to their capacity.

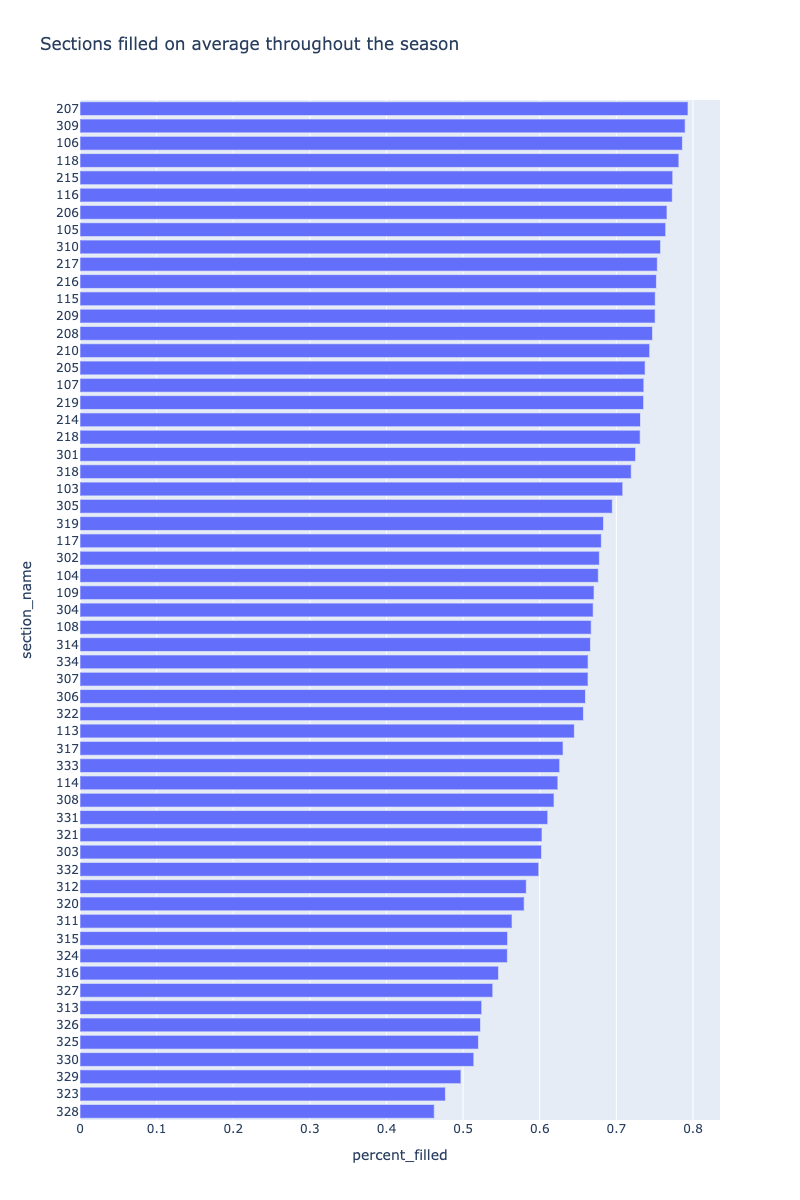
**Methodology :**

1. Grouped the SCAN dataset by event\_name and section\_name to calculate the total scans/attendees for each event for each section.
2. Merged the SEAT dataset on the above dataset. This way we have total scans and total seats in each row.
3. Calculate the percentage of seats filled for each section for each game using total\_scans/total\_seats.
4. Grouped the percentage of seats filled for each section by game.
5. Plotted these numbers onto a bar chart for better analysis.
6. Grouped the sections according to the levels and projected these ona box plot to understand the distribution of seats in Level 1, 2 and 3.

**Insight :** This plot shows the distribution of seats filled in each level. It looks like Level 2 is more consistent with around 75% attendance while Level 3 varies from 45% to ~78% of average. Seems like the whole stadium is never full but reaches its maximum of 80% attendance.



**Insight :** Here this column chart shows the capacity filled on an average for all games throughout the season. The top 10 sections filled are given below. Looks like most of the fans prefer seating on level 2.

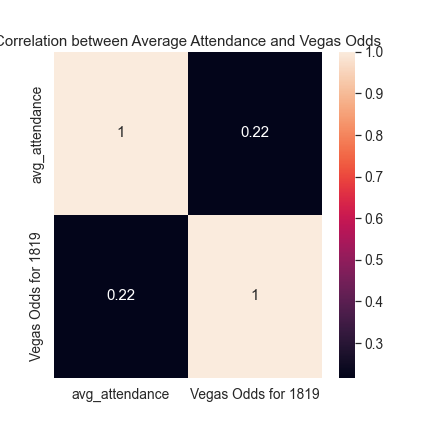


Q4) Is there a correlation between the opponent team having a higher Vegas Odds Score (indicating higher probability of winning a championship) and higher attendance?

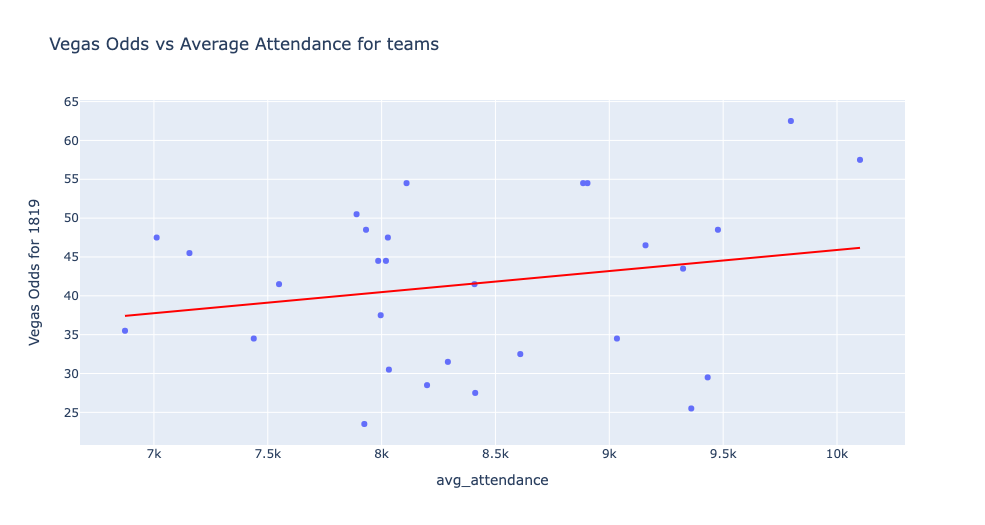
**Methodology:**

1. Using the dataset from Question 2 which contains total attendance for each game, we merge Vegas Odds from the TEAM dataset.
2. Plot a Correlation HeatMap with average attendance and Vegas Odds.
3. Use a scatter plot with a regression line for these features to visualize the spread of data.
4. With the help of a bubble chart, plot the similar scatter plot with Championships won by each team as an extra feature to check whether a franchise with more rings results in more attendance with respect to Vegas Odds.

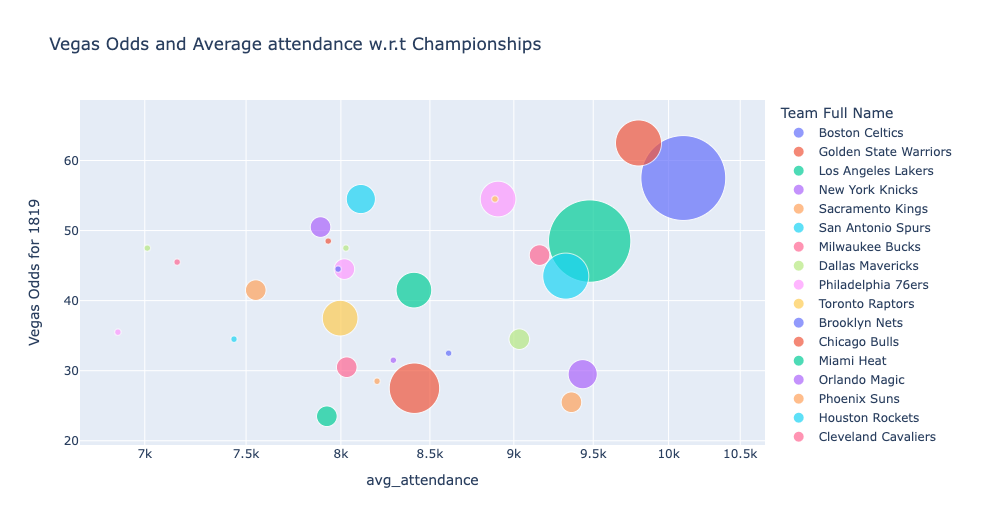
**Insight :** This Correlation Heatmap is used to compare between Average higher attendance and Vegas Odds. It seems like there is a good positive correlation between these features. People would want to watch an opponent team which has a higher probability to win for that season.



**Insight :** These graphs below show the distribution of teams to observe correlation between Average attendance and Vegas Odds. Looks like for some cases, this analogy remains true.



**Insight :** To understand this factor better, I have used the number of championships won by each team as an extra feature. From the second graph, it is safe to say that fans would prefer watching opponent teams that have more championships. For example, Boston and the LA Lakers are the teams that have a high Vegas Odds ratio as well as more championships and their correlation is highly positive.



Q5) On average, which 15-minute period before or after the start of a game has the highest number of people scanning. (Example 30 – 16 minutes prior to the start of a game)

**Methodology:**

1. Aggregated the difference between scan\_time and event\_time for each entry in the SCAN dataset.
2. Created 15 minute bins and pushed these rows onto these bins to calculate the total scans in each interval.
3. Plotted a bar chart highlighting the top 10 intervals that have the most ticket scans.

**Insight :** The bar chart below plots the 15 minute intervals on the X-axis and scan\_count on Y-axis. The yellow shade represents scans before the start of the game. The red bars represent the top 10 most intervals where fans enter the stadium. The first fifteen minutes after the start of the game is when the highest number of people scan.

