## Scenario:

I found a dataset in Big Query of pitch by pitch analysis of the 2016 MLB season. It has the type of pitch, pitch outcome, score at any given instance, attendance, batter, pitcher and fielder information. I was curious after exploring the data who won the most games and what kinds of pieces of data had a positive correlation to that.

## Strategic Goal:

**Ultimate goal of analyzing data is to determine correlation between games won and attendance or other statistical measures.**

## Results:

Through my analysis I determined that the top 5 teams in total attendance draw, all finished in the top 9 for total wins. Additionally the top team in total attendance finished top in total wins. Therefore there is a strong positive correlation between total wins and total attendance. Below is my final SQL query and further below additional explorationary queries I constructed while examining this data set

```
SELECT COUNT(game_winner) AS total_games_won,
ROUND(COUNT(game_winner)/162 * 100.0, 2) AS percent_games_won,
ROUND(SUM(attendance)/162, 2) AS average_game_attendance,
SUM(attendance) AS total_attendance,
game_winner AS team_standings,
FROM(
SELECT DISTINCT(g.gameID),g.homeTeamName, g.awayTeamName, g.homeFinalRuns,
g.awayFinalRuns,
  CASE
    WHEN g.homeFinalRuns > g.awayFinalRuns THEN g.homeTeamName
    WHEN g.awayFinalRuns > g.homeFinalRuns THEN g.awayTeamName
    ELSE 'no_winner'
  END AS game_winner,
  s.attendance
```

```
FROM mlb-2016-385422.game_data_2016.games_wide g
JOIN mlb-2016-385422.game_data_2016.schedules s ON g.gameId = s.gameId
)
GROUP BY game_winner
ORDER BY total_games_won DESC
LIMIT 10
```

I'll do this by stepping through various phases and questions in my analysis.

Questions I asked during my analysis:

1. How many unique games in each data set?
2. How many teams are reported within the data?
3. Which team scored the most runs in one game?
4. Which team(s) had the longest games on average?
5. Which team(s) had the highest total attendance?
6. Which one game had the highest attendance?
7. Which team(s) won the most games?
8. What is the average of foul balls hit in a game?
9. What is the average of ground balls hit in a game?
10. What is the average of fly balls hit in a game?
11. Does hitting more foul balls per game result in a better overall record?

Exploration details;

1. Determine the unique number of games in each data set.
    a. Games_post_wide - 28
    b. Games - 2428
    c. Scheduled - 2431

```
SELECT COUNT (DISTINCT gameId) AS unique_games_post
FROM mlb-2016-385422.game_data_2016.games_post_wide;
```

```
SELECT COUNT (DISTINCT gameID) AS unique_games
FROM mlb-2016-385422.game_data_2016.games_wide;

SELECT COUNT(DISTINCT gameID) AS unique_schedule_games
FROM mlb-2016-385422.game_data_2016.schedules
```

2. Determine a unique number of teams.
   *There are 31 unique teams*

```
SELECT DISTINCT (homeTeamName)
FROM mlb-2016-385422.game_data_2016.games_wide
```

3. Determine which team scored the most runs in one game.
   *The Blue Jays scored 17 runs in one game!*

```
SELECT homeTeamName AS highest_single_game_run_total, homeFinalRuns
FROM mlb-2016-385422.game_data_2016.games_wide
WHERE homeFinalRuns = (SELECT MAX(homeFinalRuns) FROM
mlb-2016-385422.game_data_2016.games_wide)
```

4. Which team had the longest games on average?
   *The top 3 - Arizona Diamondbacks, Boston Red Sox and Colorado Rockies - had the highest average durations of game play*

```
SELECT AVG(duration_minutes) AS average_game_duration, homeTeamName
FROM mlb-2016-385422.game_data_2016.schedules
GROUP BY homeTeamName
ORDER BY average_game_duration DESC
LIMIT 3
```

5. Which team had the highest total attendance?
   *Top 5 are; Dodgers, Cardinals, Blue Jays, Giants and Cubs*

```
SELECT SUM(attendance) AS full_season_attendance, homeTeamName
FROM mlb-2016-385422.game_data_2016.schedules
GROUP BY homeTeamName
ORDER BY full_season_attendance DESC
LIMIT 5
```

6. Which one game had the highest attendance?
   *Top 5 all were LA Dodgers home games, 4 were against the Giants and one against the Rockies*

```
SELECT MAX(attendance) AS most_attended_game, homeTeamName, awayTeamName,
startTime
FROM mlb-2016-385422.game_data_2016.schedules
GROUP BY homeTeamName, awayTeamName, startTime
ORDER BY most_attended_game DESC
LIMIT 5
```

7. Which team won the most games?
   *Top 3 Cubs 105, Nationals and Rangers both at 95 wins*

```
SELECT COUNT(game_winner) AS total_games_won, game_winner AS team_standings,
FROM(
SELECT DISTINCT(gameID),homeTeamName, awayTeamName, homeFinalRuns, awayFinalRuns,
   CASE
     WHEN homeFinalRuns > awayFinalRuns THEN homeTeamName
     WHEN awayFinalRuns > homeFinalRuns THEN awayTeamName
     ELSE 'no_winner'
   END AS game_winner
 FROM mlb-2016-385422.game_data_2016.games_wide
 )
GROUP BY game_winner
ORDER BY total_games_won DESC
```

8. What is the average number of foul balls hit during a game?
   *52 foul balls are hit on average during a game in the 2016 season*

```
SELECT AVG(foul_balls) AS foul_balls_per_game
FROM
(
 SELECT COUNT(outcomeID) AS foul_balls, gameID
 FROM mlb-2016-385422.game_data_2016.games_wide
 WHERE outcomeID = 'kF'
 GROUP BY gameID
)
```

9. What is the average of pitch speed during a game?
   * 88.14 MPH was the average pitch speed during a 2016 MLB game*

```
SELECT AVG(pitchSpeed) AS average_pitch_speed
FROM mlb-2016-385422.game_data_2016.games_wide
WHERE pitchSpeed <> 0
```

10. What are the top three pitch types thrown during the season?
    * First query I wanted to see how many distinct pitch types there were = 13*
    *Second query I averaged them out and grouped them for the year =
    Fastball 49% of the time
    Slider 15% of the time
    Curveball 10% of the time*

```
SELECT DISTINCT(pitchTypeDescription) AS average_pitch_type
FROM mlb-2016-385422.game_data_2016.games_wide
WHERE pitchTypeDescription <> ''
```

```
SELECT ROUND(COUNT(pitchTypeDescription) *100.0/SUM(COUNT(pitchTypeDescription))
OVER(), 2) AS average_pitch_type_percentage, pitchTypeDescription
FROM mlb-2016-385422.game_data_2016.games_wide
WHERE pitchTypeDescription <> ''
GROUP BY pitchTypeDescription
ORDER BY average_pitch_type_percentage DESC
```

11. Does hitting more foul balls per game result in a better overall record?
    * Top teams in foul balls per game were; Indians, Rockies, Cardinals, White Sox, Cubs, Giants, Pirates and Tigers. Of which only 2 of these teams were in the top 5 of winning records. Thus there seemed to be no strong correlation between foul balls per game and wins.*

```
SELECT team_name, ROUND(AVG(foul_balls), 2) AS foul_balls_average
```

```
FROM (
 SELECT
   CASE
     WHEN inningHalf = 'TOP' THEN awayTeamName
     WHEN inningHalf = 'BOT' THEN homeTeamName
     ELSE 'Unknown'
   END AS team_name,
   COUNT(outcomeID) AS foul_balls,
   gameId
 FROM mlb-2016-385422.game_data_2016.games_wide
 WHERE outcomeID = 'kF'
 GROUP BY gameId, team_name
) AS subquery
GROUP BY team_name
ORDER BY foul_balls_average DESC
```