

Will Pumphrey
Luis Cerrilla
John Dilligard
Cale Wallace

Super Bowl Analysis

The aim of this project was to look at the history of the Super Bowl to gain a better understanding of individual performances, team performance, Super Bowl Location and quality of gameplay. In addition, the emergence of trends since the first Super Bowl in 1967 were sought after. Kaggle had a very good, clean, and informative data set with data ranging from the first to the most recent Super Bowl with a usability rating of 10.0. Partnering the Kaggle dataset with the Google Geolocation API allowed for the information to be interpreted and analyzed.

PROBLEM STATEMENT:

This is an analysis of a dataset of the history of Super Bowls. The major purpose of this project is to answer some questions about the Super Bowls through time, such as:

- How many points do you need to win a Super Bowl?
- Who is the player with the most MVPs?
- Which teams have won the most Super Bowls?
- How many points the winners scored?
- What was the margins of victory over the years?
- Where have the Super Bowls been held?

While there were limitations with the dataset in terms of expanding on potential correlations between variables in a predictive manner, the initial questions were all answered through the research and analysis.

DATA:

The data used was selected because it was clear, usable, from a reliable source, and had pertinent information. The usability was a 10.0 and while it contained only 54 rows, the information was useful in creating an image of different correlations and an interesting trend in the Super Bowl.

| | Date | SB | Winner | Winner Pts | Loser | Loser Pts | MVP | Stadium | City | State |
|---|------------|-------------|----------------------|------------|----------------------|-----------|-----------------|-------------------------------|-----------------|------------|
| 0 | Feb 2 2020 | LIV (54) | Kansas City Chiefs | 31 | San Francisco 49ers | 20 | Patrick Mahomes | Hard Rock Stadium | Miami Gardens | Florida |
| 1 | Feb 3 2019 | LIII (53) | New England Patriots | 13 | Los Angeles Rams | 3 | Julian Edelman | Mercedes-Benz Stadium | Atlanta | Georgia |
| 2 | Feb 4 2018 | LII (52) | Philadelphia Eagles | 41 | New England Patriots | 33 | Nick Foles | U.S. Bank Stadium | Minneapolis | Minnesota |
| 3 | Feb 5 2017 | LI (51) | New England Patriots | 34 | Atlanta Falcons | 28 | Tom Brady | NRG Stadium | Houston | Texas |
| 4 | Feb 7 2016 | 50 | Denver Broncos | 24 | Carolina Panthers | 10 | Von Miller | Levi's Stadium | Santa Clara | California |
| 5 | Feb 1 2015 | XLIX (49) | New England Patriots | 28 | Seattle Seahawks | 24 | Tom Brady | University of Phoenix Stadium | Glendale | Arizona |
| 6 | Feb 2 2014 | XLVIII (48) | Seattle Seahawks | 43 | Denver Broncos | 8 | Malcolm Smith | MetLife Stadium | East Rutherford | New Jersey |

Super Bowl Analysis

DATA CLEANING:

The data came from a CSV, so that file was read in so the data could be cleaned and analyzed. The data was relatively clean and did not have missing or misplaced values and was void of errors. However, there were a few small things that needed to be changed for enhanced usability. One thing that needed to be changed was the date (it was in Feb-1-2000 format). The data set had the date and it was changed to date time using `df.pd_DateTime`. The stadium locations were given without geolocational coordinates, so those had to be researched further. A new column was created titled "Margin of Victory".

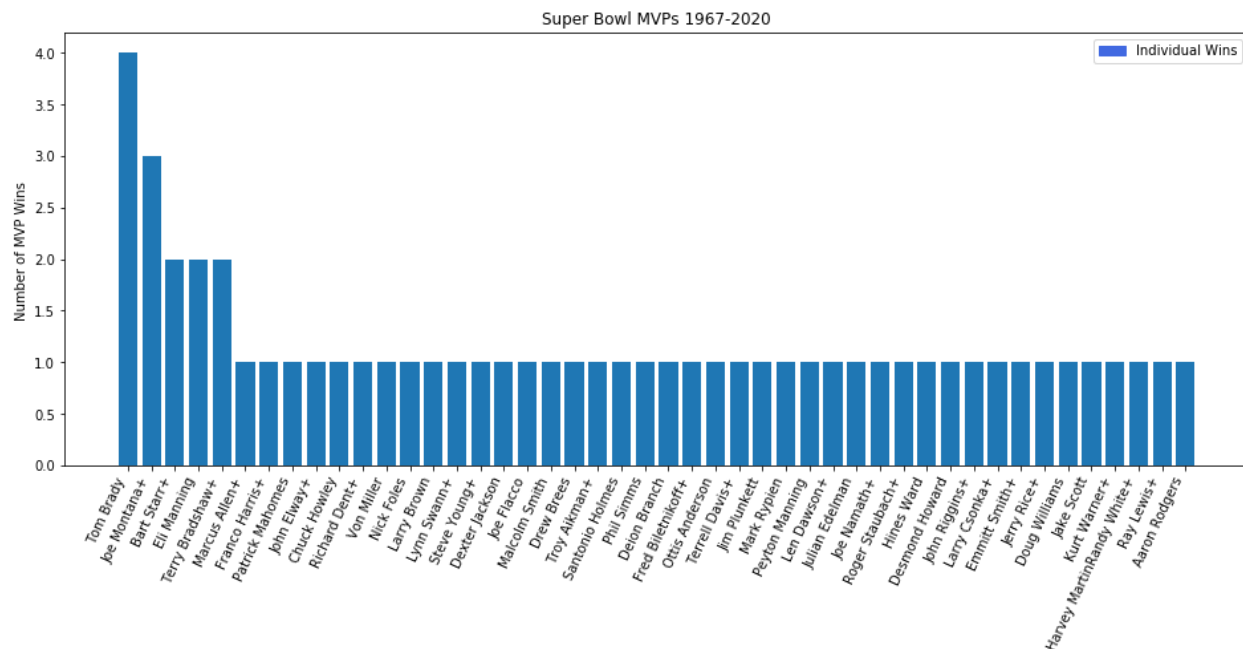
ANALYSIS: How many points do you need?

The following data frame was used to create a display table of the winners' average points, losers' average points, and the difference in the total number of points scored through the years. The total through the years was used to show how the number of points scored by the winner was almost double the points scored by the loser.

| | Winner Pts | Loser Pts | Overall Point Differential |
|---|------------|-----------|----------------------------|
| 0 | 30.11 | 16.2 | 751 |

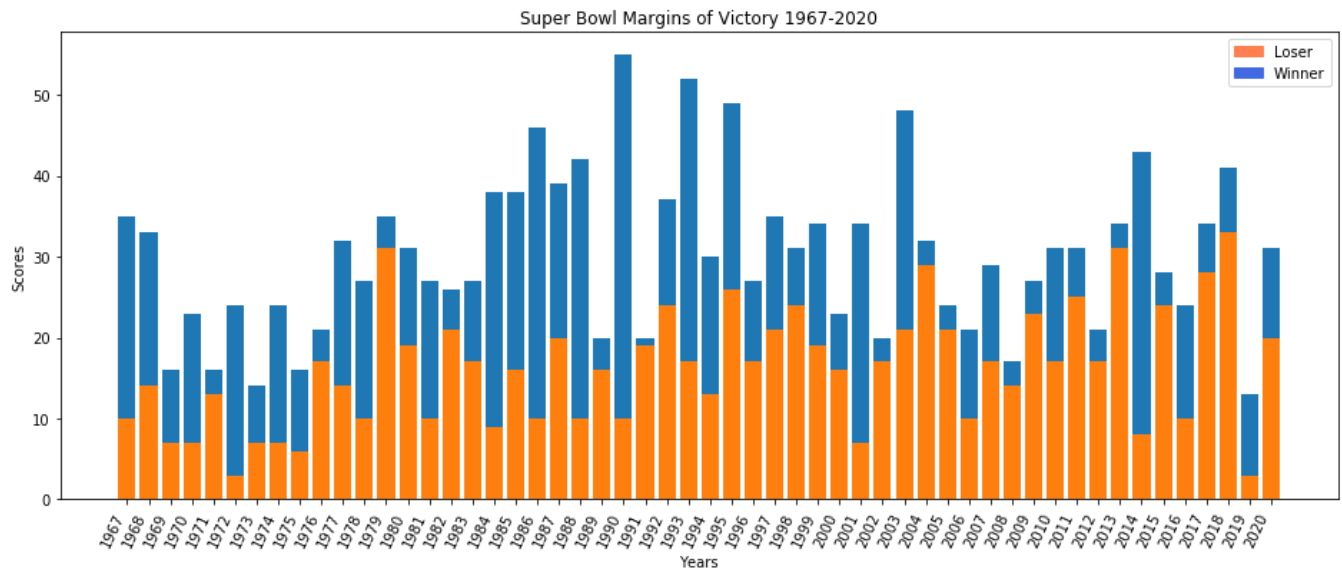
ANALYSIS: Who has the most MVPs?

Tom Brady has won the most MVPs.



Super Bowl Analysis

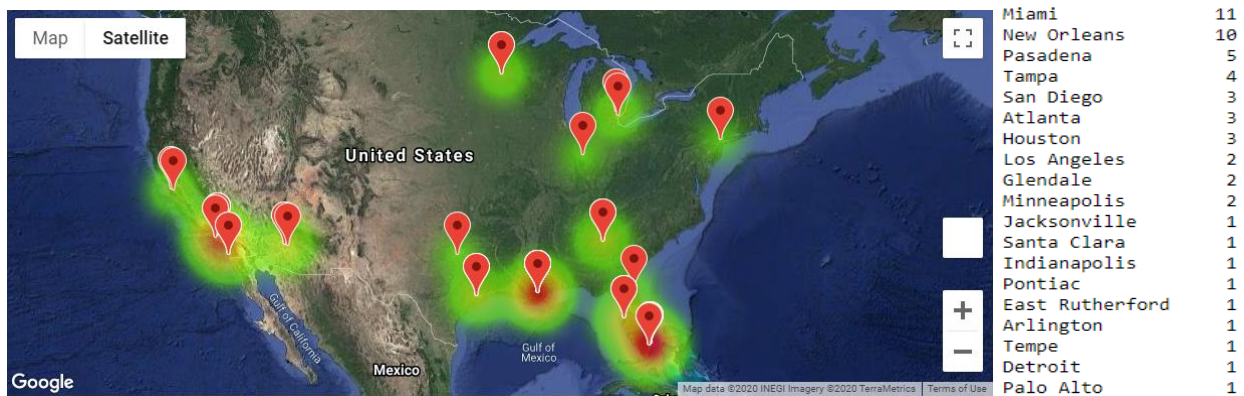
ANALYSIS: What is the margin of victory?



The margin of victory has decreased in the last few years, with one major exception (2014). The games have been more competitive in recent years.

ANALYSIS: Where have the games been played?

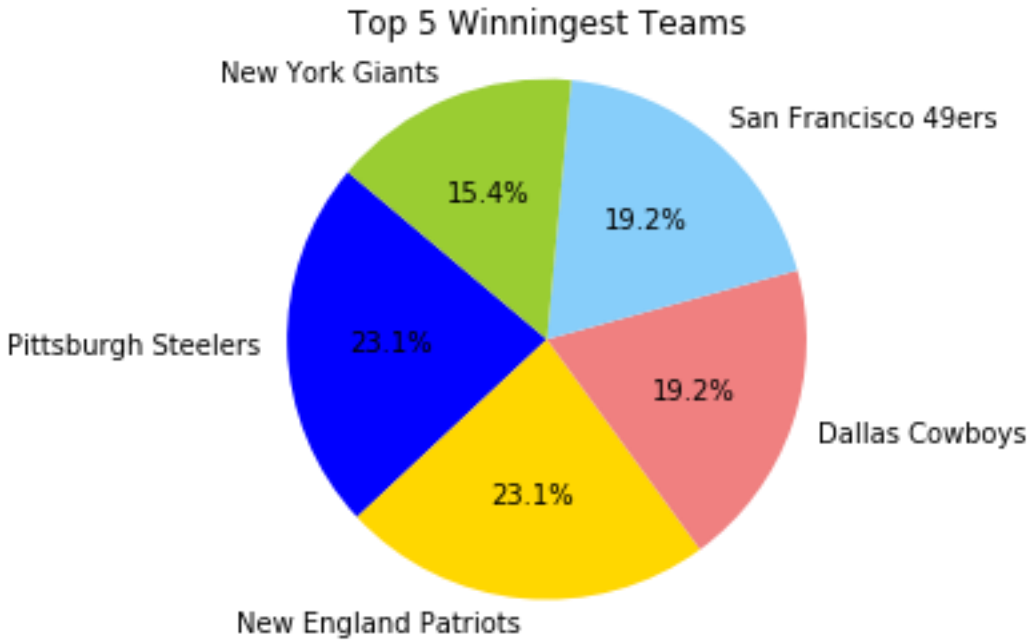
A heat map of the US was created with markers showing the location and frequency of Super Bowl locations. The most frequent locations were found to be Miami (11) and New Orleans (10).



Super Bowl Analysis

ANALYSIS: Which teams have won the most?

Six teams have won 26/54 Super Bowl titles. The top 5 displayed below by percentages won of those 26:



Super Bowl Analysis

CONCLUSIONS:

To answer the questions in the order in which they were asked:

- **How many points do you need to win a Super Bowl?**
 - It was concluded that the number of points to win was 17 points based on the average points scored by the losers, but the points scored by the winners stayed consistent at about 32. The trend with the point differentials and margin of victory over the years are affected by the same skew. The skew comes from the fact that there has been a considerable change in recent Super Bowls.
- **Who is the player with the most MVPs?**
 - Tom Brady has won the most Super Bowl MVPs with 4. Joe Montana finished second with 3.
- **Which teams have won the most Super Bowls?**
 - The Patriots and the Steelers are tied for the most Super Bowl wins of all-time with 6 per team.
- **How many points the winners scored?**
 - While the points scored by the loser has increased in recent years, the points scored by the winner has stayed about the same at around 32 on average.
- **What was the margins of victory over the years?**
 - It was reasonable to conclude that while the points scored by the winning team stayed fairly consistent over time, with a few outliers, the points scored by the losing team has actually increased in the last few Super Bowls, meaning that the games are becoming more competitive. The limitation here is that there are so many factors at work when it comes to why this phenomenon is happening. Potential reasons for this are the earlier games were played between two leagues that had recently formed into one, and one may have had a competitive edge; contract structures of the league/players; overall conditioning of the players.
- **Where have the Super Bowls been held?**
 - Miami was first with 11. New Orleans was second with 10. Los Angeles was third with 7. It was reasonable to conclude that since the majority of the games were not played in major market cities like New York or Chicago, and per the NFL, Miami and New Orleans were and continue being selected because of the warm weather in the winter and capabilities of the cities to play host (i.e., hotel availability).