

**THE HISTORY OF SUPER BOWL ANALYSIS IN  
RELATION TO DIFFERENT VARIABLES  
(1967-2024)**

**Ganta Supriya [ Data Visualization]**

**Part 3: Final Project**

## Introduction and Background

The history of SuperBowl is an interesting topic to gain extra knowledge upon because of the name and fame of it specially in America followed by people all around the globe. The topic SuperBowl was chosen in context to initiate the project and research about the World-famous American sporting event, SuperBowl. The selection of the topic, including the decision to present visualizations on the given topic was due to the immense interest of all the group members in the given topic itself. SuperBowl is a very famous game more like a festival and ritual for everyone in the United States of America. Because of the craze and love of the people towards the game, it became a topic of interest for our group.

The SuperBowl stands as a peak interest and is valued as the top game for the American sports culture. The topic has a lot of points to discuss and present visualizations for. The data for Super Bowl consists of various variables like the State the game was played in, the winners, the number of times a certain team has won, the number of times certain player has been awarded the Most Valued Player, the points that a certain winner has received in compared to the losing points and many more. These many options for the data drove our interest in doing further research on the given topic.

The main aim of the project is to discover various relationships between the wins and the loss of the games and in association with different factors like players, city the game is played in, state the game was played in and many more. The data over the years gives the viewers and the analysts a major idea of the factors affecting the win and the loss and what and how the game is played and won.

The project is a compilation of different visualizations and dashboards for the viewers to analyze and view the project from different perspectives. The project gives a detailed analysis about the players of the Super Bowl, their wins, their loss, the state they present in a certain game, the number of times the players won and lost and many other relevant information which is useful in performing analysis through detailed visualizations of different types including maps, bar graph, pie chart, line chart and many more. Because of the presence of multiple data sources and data combination, it is easy for the analysts to generate discrete charts and present the findings.

Through the project, we have tried to figure out the relation among various factors influencing the game status whether it's a win or a loss and various outcomes from the points differences.

This project and the topic is useful to people who are very much involved in the game including the player himself, the coaches who train the players, the team as a whole, the people who bet for different games, the investors who are investing in the games and the teams and all in all the fans and the viewers of the games who are interested in the game. The players can benefit because the project and the visualizations help the player make an analysis on what their strengths and weaknesses are which is also helpful to the coaches and the investors of the game so the overall improvement can be made in the future.

## Data Set

The SuperBowl dataset comes from the most popular American sporting ritual and a worldwide known game SuperBowl played in the United States which is also the biggest known yearly league championship of the National Football League (NFL). The datasets utilized in this project come from as old time as 1967 range till date.

Caroline Arnold is the curator of the dataset (Visual Data Science). It is reliable and suitable for analytical and instructional uses because it is publicly available on Kaggle and OpenDataSoft.

This dataset is based on actual data from 1967 to 2024. The Super Bowl dataset facilitates the analysis of every team player's statistic and performance in relation to many factors, such as stadium, victory margin, points differential, etc. The data used for this project is also real time data.

This dataset includes statistics on every Super Bowl game played between 1967 and 2024, including information on the teams, attendance, point differential, and winners and losers, as well.

as officials and MVPs. The data is collected manually and organized to present variations of graphs and visualizations. Because the dataset is manually done the data cleaning is not necessary.

## Data Columns Description

Date	The Super Bowl game's date. (Time)
SB	The Super Bowl game's number. (Number)
Attendance	The total number of spectators at the match. (Number)
QB Winner	The game's victorious quarterback. (String)
Coach Winner	The squad's victorious coach. (String)
Winning Pts	The amount of points the victorious team scored. (Number)
QB Loser (String)	The quarterback who loses the game. (String)
Coach Loser	The team that lost the match as a coach. (String)
Losing Pts	The amount of points the losing team managed to score. (Number)

MVP	The most valuable player of the game is Mvp. (String)
Stadium	The location of the game's play. (String)
City	The location of the game's play. (String)
State	The virtual state of the game. (String)
Point differential	The number of points that separates the victorious and defeated teams. (Number)
QB player Jersey Number	The number of jerseys of each player (Number)
Player city	The city that the player was born in and represent in each game (String)
Player State	State the player was born in and represent in the game (String)
Viewership Rating	The ratings provided by the viewers in accordance with the game who watch the game through TV and other media (Number)
No of TV Viewers	The number of TVs the show is being watched from all around the world. (Number)

The dataset has unique values, and its size allows users for the creation of several visualizations and analysis through various combinations, comparisons, and relationships. Indeed, the generation of a timeline can also be done with the help of the dataset that spans over the years from 1967 through 2024. The dataset is a perfect set to fulfill the project objective which is to analyze the game and its affecting factors.

## Data Stories

### Chart 1

#### Which state has the greatest number of wins in the history of SuperBowl?

This chart shows the number of SuperBowl wins by state. The variables used to generate the visualization are-

1. Longitude and Latitude in columns and rows which help generate the map.
2. Details are shown by State in the marks card.
3. Count of Date column added in the color section of marks card.
  - CNT: This stands for "Count." A Tableau aggregation function that calculates the total number of occurrences for a field within a specific context.
  - (Date): This specifies the field you want to count. In this case, it's the "Date" field.

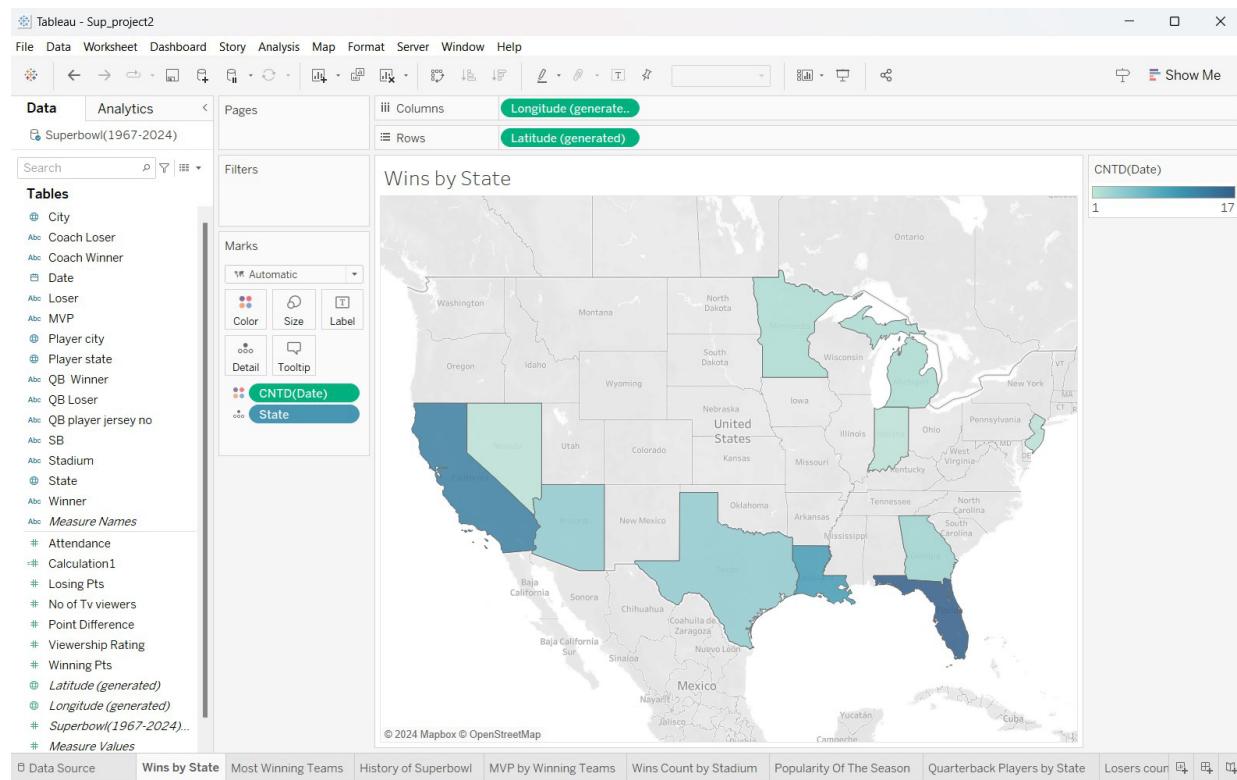


Fig1- SuperBowl Wins by State

The Super Bowl map visualization shows number of wins across the U.S. states. California and Florida are on the top with 17 and 13 wins respectively. This chart shows more than just the

number of wins; it tells a story about the love for the game and the sense of community that the Super Bowl game represents for the country.

## Chart 2

### **Who has the most popular MVP presence in the SuperBowl leagues?**

The following visualization tells us about the MVP's presence by the winning teams. The data is sorted by both winning points and the sum of attendance for each MVP's corresponding Super Bowl victory. The bar chart depicts the stadium attendance and winning points associated with each MVP.

The variables used to generate this chart are:

1. MVP in columns. This shows the MVP names in the column generating the x axis of bar graph.
2. SUM (winning points) for each MVP across the y-axis in Rows. This shows the winning points scored when this player was in a SuperBowl winning game.
3. Sum (Attendance) is added to the color in the marks card. The darker shade of blue represents the high stadium audience.

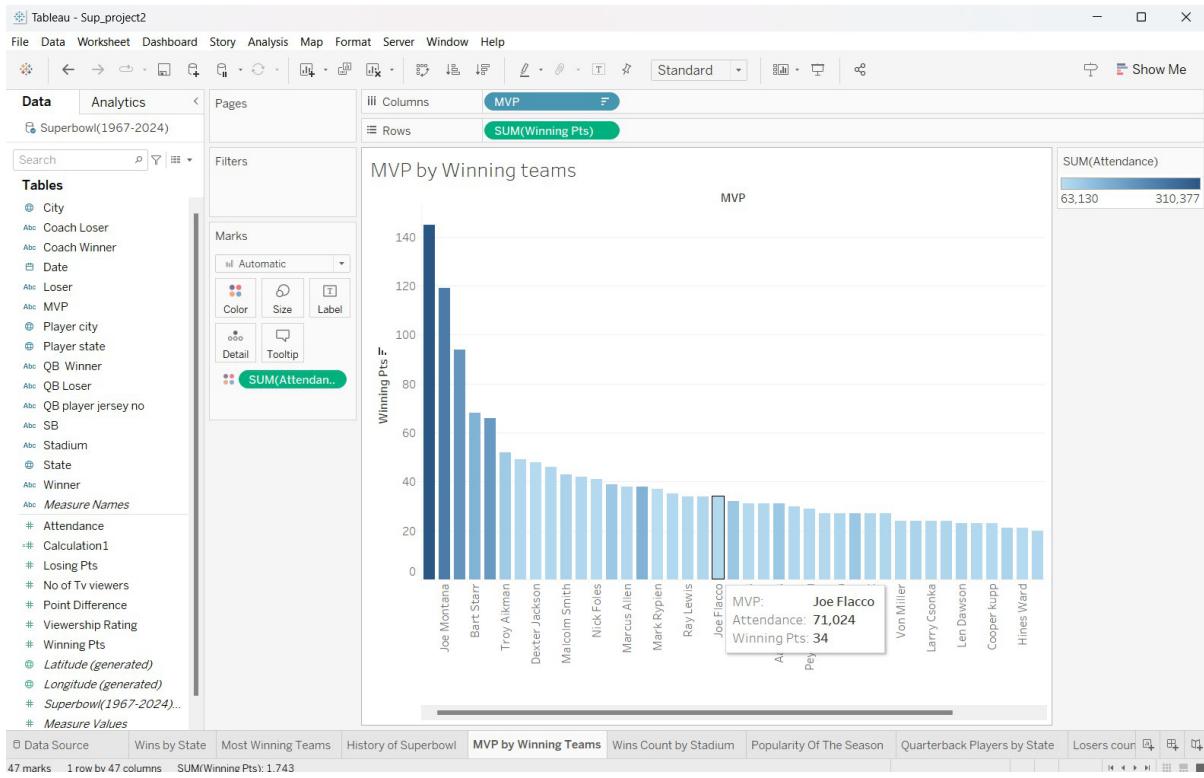


Fig2- MVP by Winning Teams

This visualization delves into the history of Super Bowl winning points, highlighting Tom Brady as the Most Valuable Player (MVP) with the highest combined score for both winning points and total attendance. It suggests a potential correlation between his presence and both fan engagement (attendance) and team success (winning points) followed by Joe Montana and Patrick Mahomes.

### Chart 3

#### Which franchise has triumphed the most?

The visualization represents the number of times a particular team has won the SuperBowl title.

As this is a bubble chart, the size of the bubble chart corresponds to the count of the times won.

Variables and features used to create this chart are:

1. CNT(Winner) in colors option of marks card which is diverging Orange Blue. This indicates that the team which has won most times is in the shades of blue.
  2. CNT(Winnter) in the size option of the marks card. The higher the number of wins for the team, the larger is the bubble. Which makes the visualization easy to read at a glance.
  3. The winner in the text option of marks card helps us show the team's name on the visualization. Each bubble represents one team based on color and size.

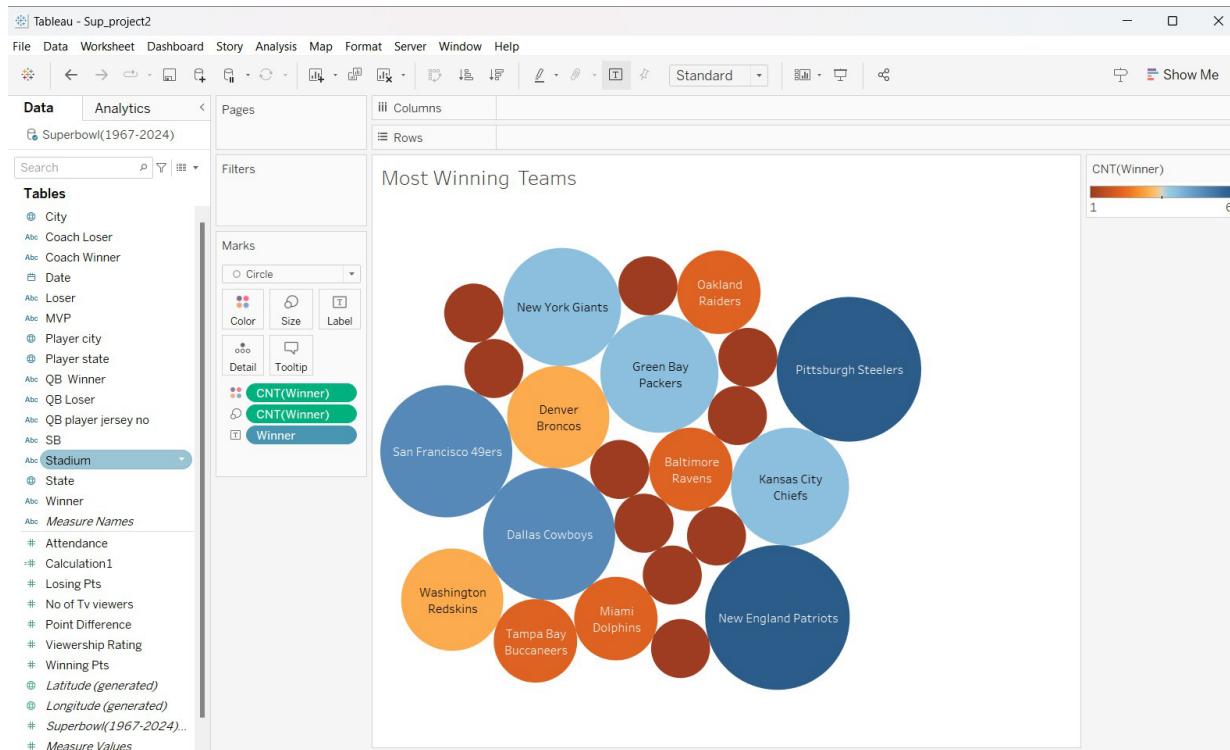


Fig3- Most Winning Teams

The visualization tells the number of times a team won the title of SuperBowl. Clearly New England Patriots and Pittsburgh Steelers have won the title maximum number of times i.e. 6 followed by San Francisco 49ers and Dallas Cowboys at 5 wins each.

## **Chart 4**

**Display the winner and the runner up for each SuperBowl season.**

This stacked bar graph represents the story of the winner and runner up by points for each season of SuperBowl from 1967 through 2024. The green part in the graph represents the winning points gained by the winning team and the orange part represents the points scored by the runner up team in each respective SuperBowl.

The graphical break down-

1. Year (Date) and Winner in the Rows represents the Year and the winning team by the year.
2. The Measure Values are SUM (Losing Pts) and SUM (Winning Pts) which are then dragged across the marks card for representing the points by size and text.
3. The details represent- the name of winning and losing team and respective points with year of the match.

Analyzing the stacked bars, we can observe variations in the point differential between winners and runners-up. Some Super Bowls were close games, with bars showcasing a near-even split between green and orange. However, other Super Bowls witnessed dominant victories, with the green segment towering over the orange, indicating a significant point disparity.

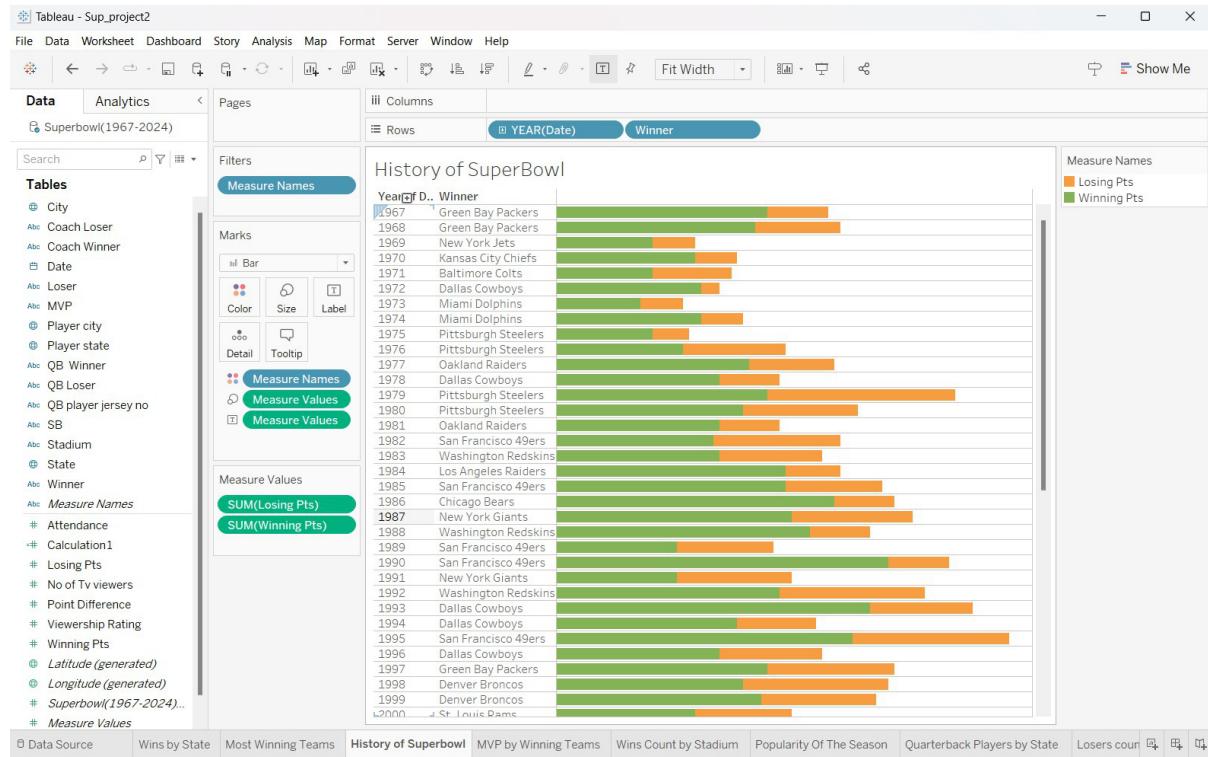


Fig4- History of SuperBowl

This data can be represented using a different visualization in the form of trend chart for winning and losing points across each season.

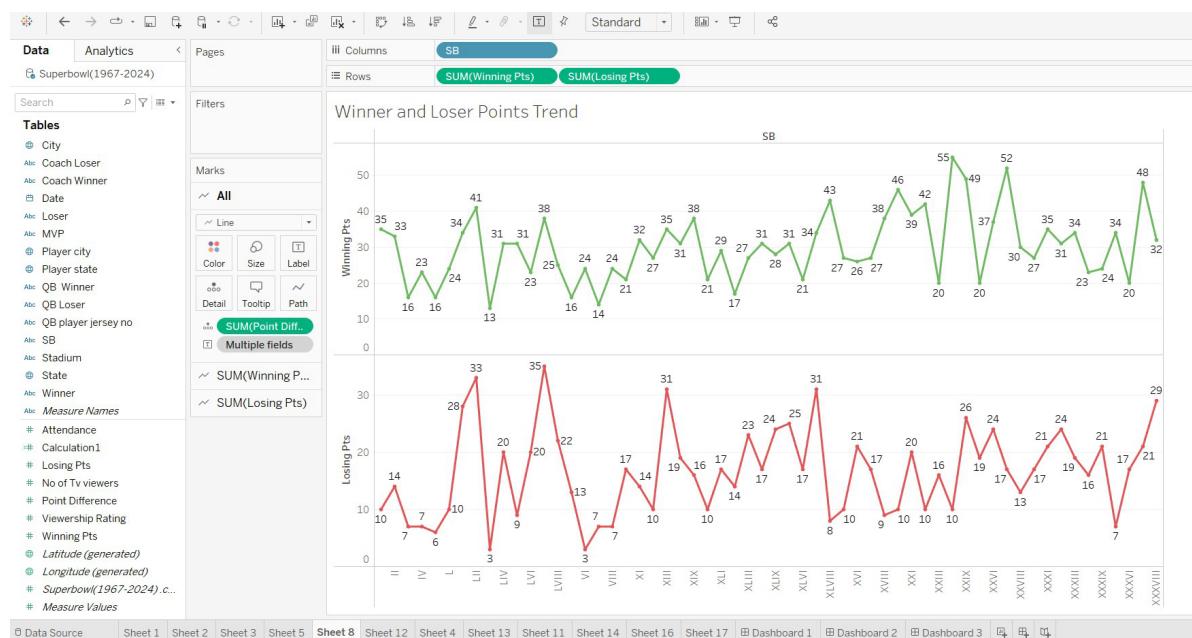


Fig4.1-History of SuperBowl

Further the analysis includes –

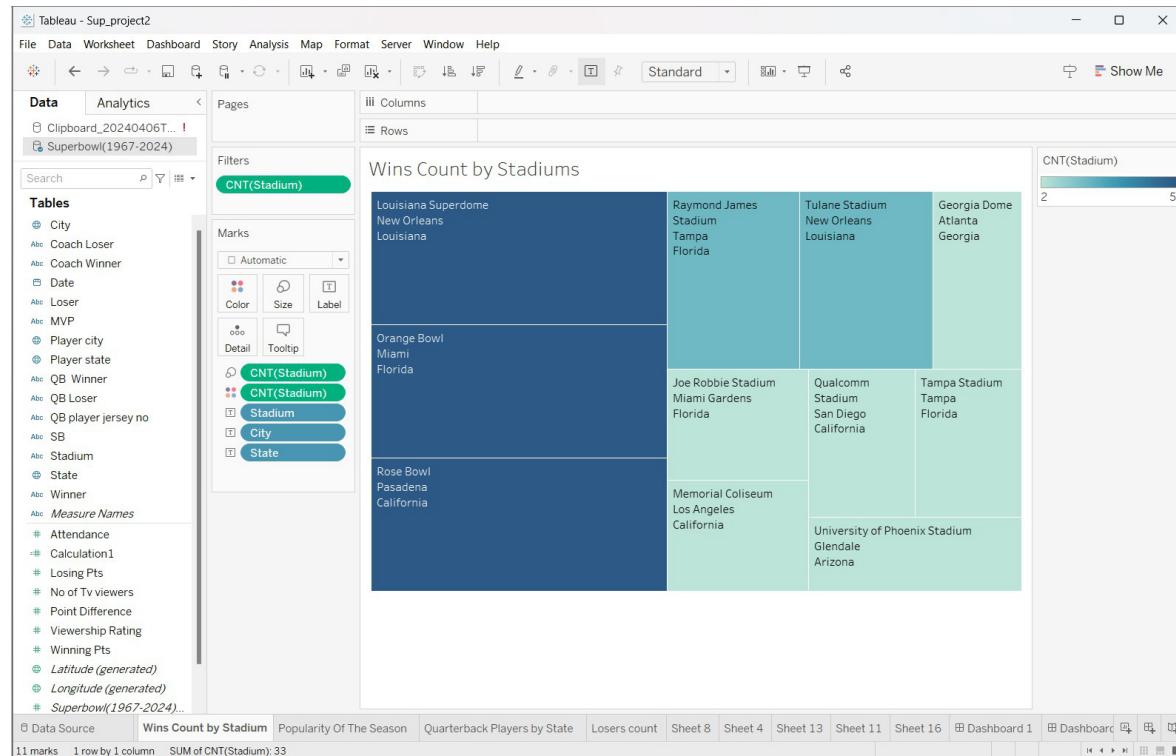
Examining the distribution of point differentials may give information about the overall competitiveness of the Super Bowl. The 57<sup>th</sup> SuperBowl game as we can see in the graph above (LVII) was a close call i.e. 35 losing points – 38 winning points which was a 3-point difference in the game.

## Chart 5

### Which is the most frequent SuperBowl host stadium?

This visualization, called a heatmap, uses a clever color scheme to reveal the distribution of Super Bowl games across different stadiums. The intensity of the color – darker shades signifying a greater number of events and lighter shades signifying fewer – allows for a quick visual understanding of which venues have hosted the most championship games.

The CNT(Stadium) counts the number of games held in the stadium. The details represented are Stadium, City and State with CNT(Stadium) in the size and color option of the marks card.



### Fig5- Game Count by Stadiums

The color distribution of Super Bowl stadiums indicates their frequency. Darker squares indicate the most frequent hosts, while lighter ones indicate infrequent or non-hosting venues.

Geographic patterns may also be observed, suggesting a concentration of Super Bowls in specific regions.

### Chart 6

#### A Look at Super Bowl Audience Engagement: Which were the Peak Seasons?

The data below represents three charts for the measure of audience engagement through the years of SuperBowl.

The **first chart** shows SUM (attendance) which represents the number of fans attending the stadium game.

The **second chart** displays the SUM (No of Tv viewers) which corresponds to the engagement through the tv viewing of the game.

The **third chart** depicts the viewership rating, expressed as a percentage SUM (No of viewership rating).

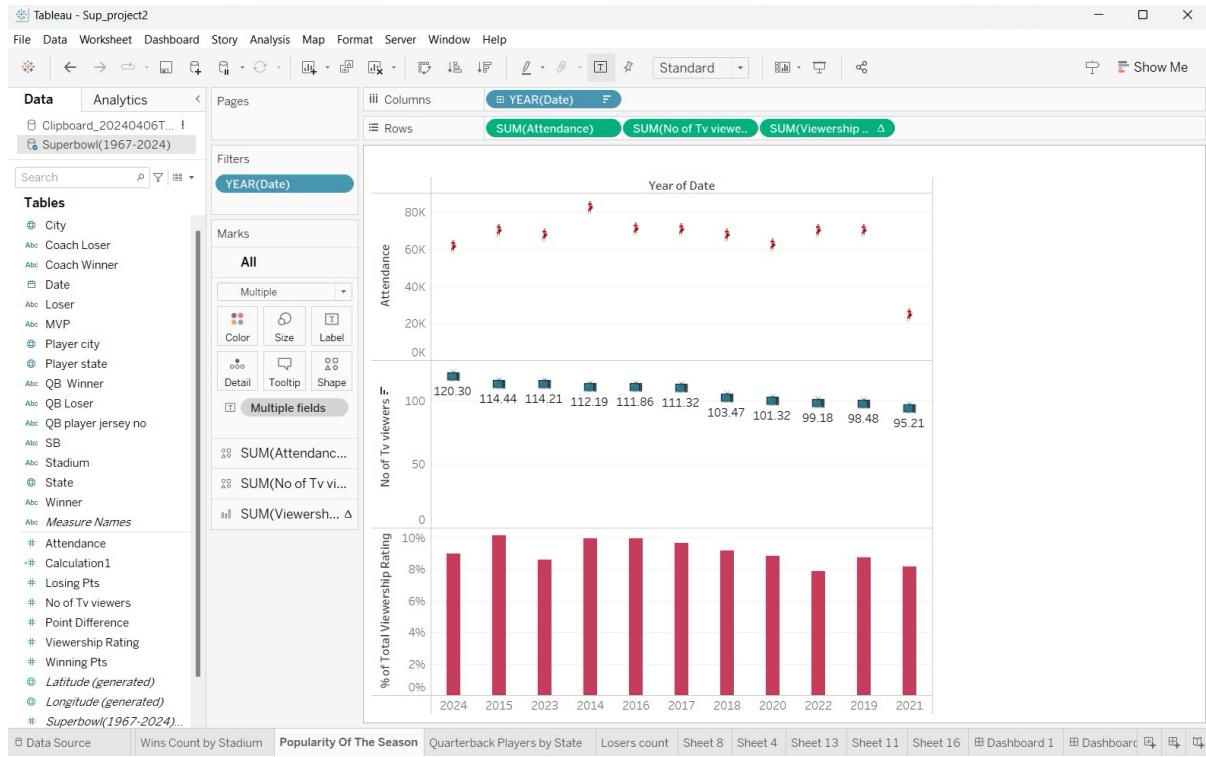


Fig6- Popularity of the Season

This graph dives into Super Bowl viewership trends, showcasing the most popular games based on television audience. It sorts viewership numbers by year, revealing Super Bowl LVIII (2024) as the most-watched game in television history. Interestingly, the graph also displays stadium attendance for each year, with Super Bowl XLVIII (2014) boasting the highest in-person attendance.

## Chart 7

### Top Quarterback Factory: Which State Reigns Supreme?

The following geographical chart represents the count of players by state.

The CNT (player state) represents the count of players belonging to each state by diverging Green-Red color. Red represents the state with the maximum number of players.

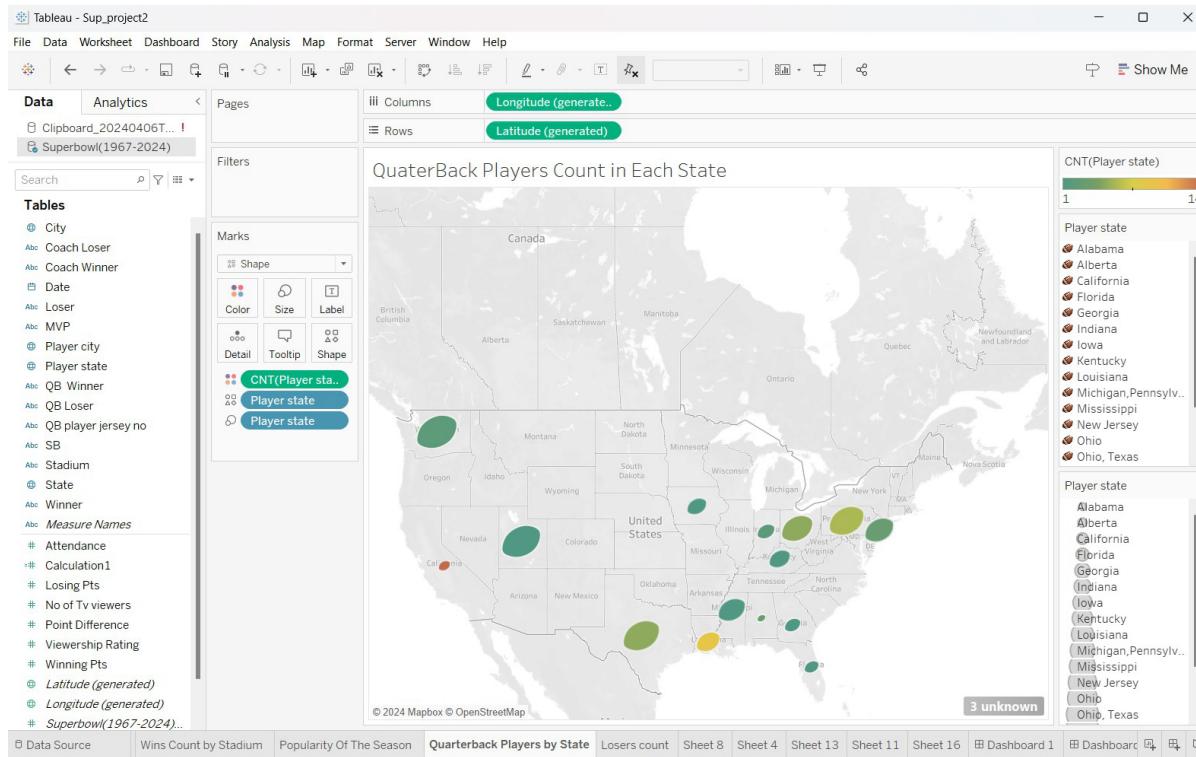


Fig7- Quarterback Players by State

Quarterbacks across America: This chart unveils a geographically diverse landscape of quarterback production. California reigns supreme with 14 players, but Louisiana puts up a strong show with 9, highlighting the talent spread across the nation.

## Chart 8

Which Winning Team attracted the largest crowd in the stadium?

The graph shows a line chart of trends between the winning teams and the attendance in the stadium during the match. The color represents the winning teams by SUM (attendance) in the label tool of marks chart. We have also added the filters by year to check the attendance by specific year if required.

The variables used to create the chart are:

1. Winner and SUM (attendance) in the columns and rows respectively.
2. Winner in the color and SUM (attendance) in the label section of the marks card.

### 3. Filter – Year (Date) represents the Year filter.

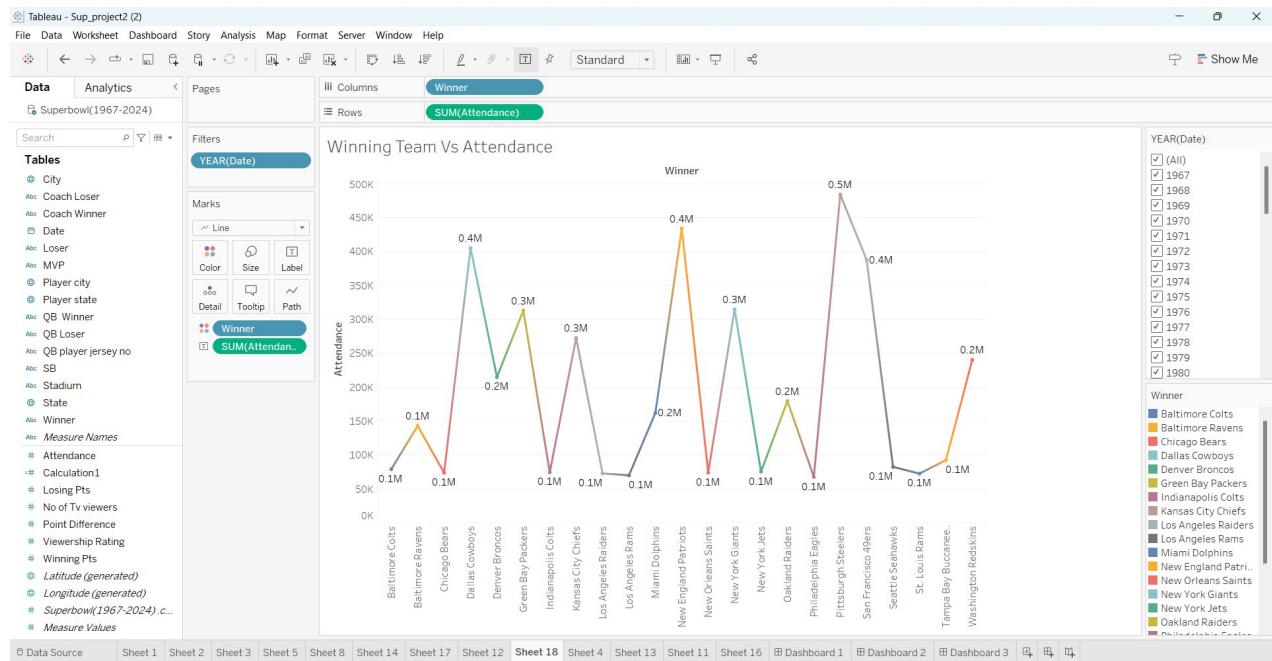


Fig8- Winning Team Vs Attendance

From the chart we can see that the San Francisco 49ers have highest attendance record cumulatively with 0.5M in stadium attendance. Followed by Dallas Cowboys, New England Patriots and Seattle Seahawks tallied at 0.4M.

### Chart 9

This dashboard offers a detailed look at the league's Most Valuable Players (MVPs). It uses a range of chart options to highlight the top five MVPs and their accomplishments.

**First chart-** A Bubble Chart shows the MVP with the most championships, with bubble size proportional to the amount of MVP awards each player has earned.

**Second chart-** A Bar Graph shows how many victories an MVP has, making it easier to compare their total success.

**Geographical map-** A final visualization examines the geographical distribution of these MVPs, identifying the states from whence they come.

The geographical chart gives a better understanding of the Top 5 MVP visualization when compared to the first two charts.

In conclusion, This MVP dashboard, with its many visualizations, provides a starting point for investigating player accomplishments, team performance, and probable geographical variables influencing the growth of outstanding NFL quarterbacks.

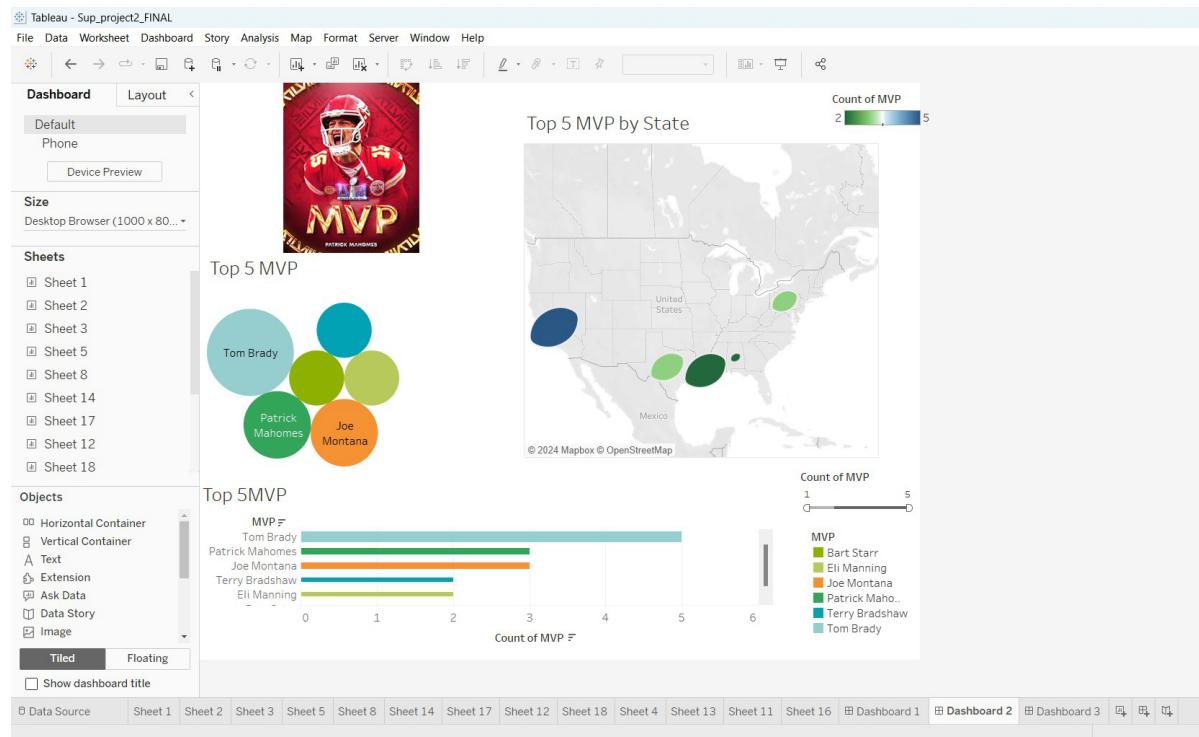


Fig9- Top 5 MVP

## Dashboard -1

### Unveil the Legacy of the Super Bowl: A 58-Year Journey

The dashboard deep dives into different aspects of SuperBowl game which answers several questions such-

1. Who are the top 5 MVPs of all time in SuperBowl history?
2. Which state produced the greatest number of Quarterback players?
3. Which game was the most competitive?

This dashboard examines NFL quarterback performance by team, providing a thorough overview of their achievements. Surprisingly, the stats show California as a top generator of quarterbacks, with the renowned Tom Brady, the league's most decorated MVP, hailing from the Golden State. Texas is close behind, with five quarterbacks in the league, including Patrick Mahomes.



Fig10-Dashboard1

In Conclusion,

This dashboard looks beneath the surface, showing the location of quarterback talent and the great characters who have left their mark on Super Bowl history. Take a 58-year journey and uncover the tales that define Super Bowl glory!

Quarterbacks: To see which states consistently produce the most quarterbacks.

Evolving Victories: Trace the changing landscape of winning margins over time with a graph showcasing Super Bowl champions and their point totals.

MVP Spotlight: Celebrate the top 5 MVP players who orchestrated championship runs throughout the SuperBowl.

## Dashboard 2 –

### Super Bowl Stories: Venues, Fans, and Champions

Forget the final score! This interactive dashboard uses data visualizations to guide you through the rich history of the Super Bowl. This dashboard includes a heatmap that depicts the geography of Super Bowl excellence, highlighting the stadiums that have hosted the most championship games. The following graph illustrates the development of television viewership, with Super Bowl LVIII (2024) being the most-watched game ever. Shifting gears, two additional graphs unveil trends on the field. One tracks team win counts, highlighting the most consistent championship contenders. The other explores points scored, revealing how offensive dominance and defensive strategies have evolved over time.

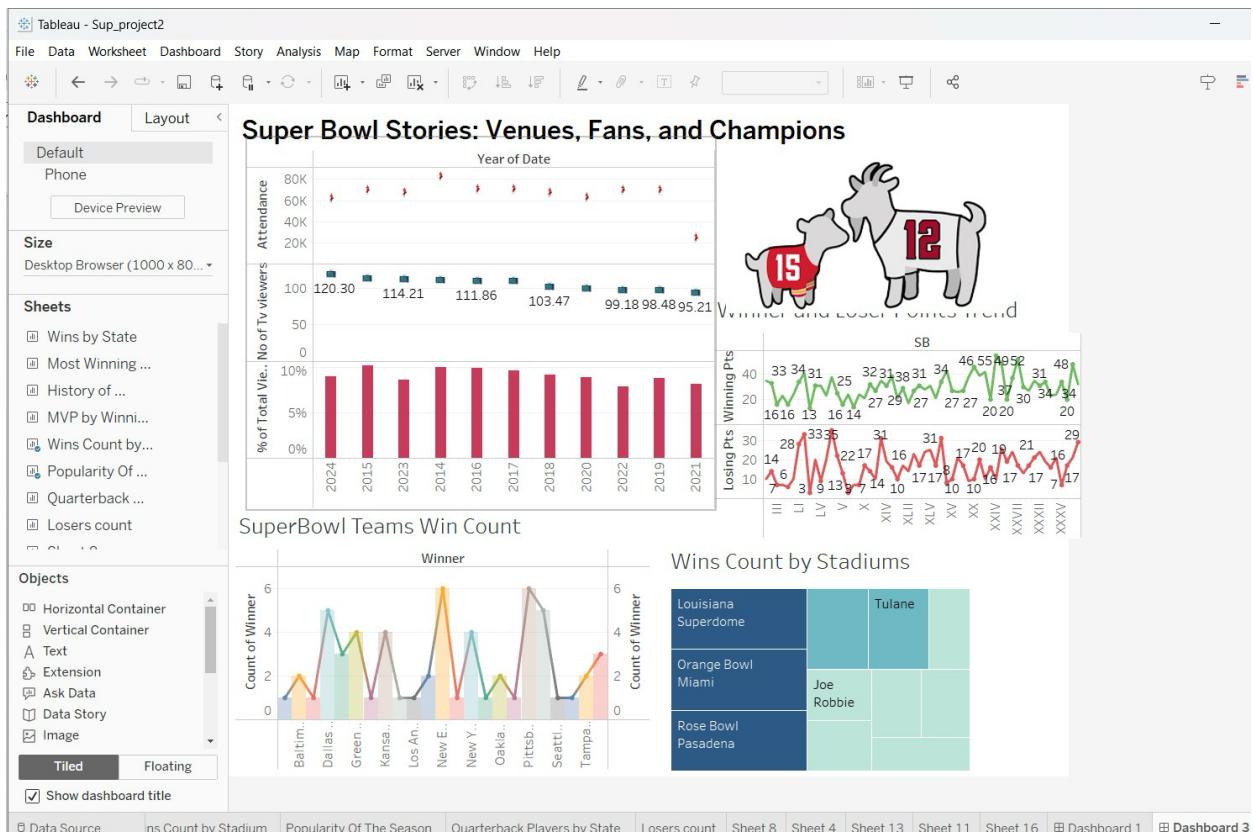


Fig11-Dashboard2

## **Summary and Conclusion:**

From the visualizations and the graphs, we can summarize that the SuperBowl has different factors which determine the wins and loss of the teams. The trends and the graphs are simultaneously formed to make the relationship more distinct and to give a clear picture of our findings.

We have been able to calculate things like Most Valued Player by Winning Teams where we analyzed the attendance and winning points; likewise, we also calculated the greatest number of winning teams; winning and losing teams over the years in consideration with different teams; the trend of winners and losers and many more. Through the dashboard, it came to our knowledge that the winners could be decided in relation to different factors such as the state, the time, the count of players and accordingly the MVPs also got their trends and analytics. This helps the viewers and analysts along with players and coaches gain a brief knowledge about their weaknesses and strengths and where and how can the games be improved and won. The dashboard is also useful to see the number of win count of the SuperBowl as per different teams, the win count as per the stadium played in and the pattern followed by the winning and the losing teams which also helps the viewers to know that a certain win or loss is not solely because of one single reason but it is due to various reasons that persists in a match.

These visualizations helped us to understand the importance of proper data presentation in relevance to understanding the context and relationship of a certain data in an appropriate format. The presentation of data is important in a high-level organizational module to create the decision makers of the organization to gain an unbiased perspective for their decisions and it also helps to critically analyze all the given data and information. The dataset is manually collected from various articles and data sources which makes the data organic and clean without any biases throughout the history of the game.

In conclusion, the project is a great help to the team members as well as the viewers to gain a deep insight about the historical football game of the United States, the SuperBowl because the project has articulated the relationship of the game in terms with various factors that could affect the win and the loss of the teams. This project is also of use for great analysis and projections for future games. It provides the players and the teams with the place for their improvement as well as to make the games better in the future through the help of ratings. Overall, this project has

summarized the importance of visualizations and proven how practical visualizations through the help of tableau can make the operations in an organization smoother and the decision-making process faster and more accurate.

## References:

1. TiMoBoz. (2020). Super Bowl History, 1967-2020. Kaggle. [Online]. Available: [“https://www.kaggle.com/datasets/timoboz/superbowl-history-1967-2020”](https://www.kaggle.com/datasets/timoboz/superbowl-history-1967-2020)
2. Tayyab Ali. (2019). The Super Bowl Dataset. Medium. [Online]. Available: <https://medium.com/visual-data-science/the-super-bowl-dataset-185e4b965dfb>
3. Tayyab Ali. (2020). NFL Super Bowl Historical Data Visualization EDA. Kaggle. [Online]. Available: <https://www.kaggle.com/code/tayyabali55/nfl-super-bowl-historical-data-visualization-eda>
4. Sports Media Watch. (n.d.). Super Bowl Ratings – Historical Viewership (1967-2020). [Online]. Available: <https://www.sportsmediawatch.com/super-bowl-ratings-historical-viewership-chart-cbs-nbc-fox-abc/>
5. Statista. (2021). Super Bowl - MVPs. [Online]. Available: <https://www.statista.com/statistics/1202199/super-bowl-mvp/>
6. Meghana P. (2020). Analyzing the Superbowl History Dataset (1967-2020). Towards Data Science. [Online]. Available: <https://towardsdatascience.com/analyzing-the-superbowl-history-dataset-1967-2020-fdee01a760c9>

