

**TASK**

**Exploratory Data Analysis on the NFL TV, Halftime Shows and the Big Game Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The National Football League Super Bowl is a huge sporting spectacle in the Unites States of America. This spectacle is known for its crazy comebacks, expensive adverts and the nail-biting suspense and drama. The NFL Super Bowl halftime shows also normally have the biggest musicians in the world performing. In this dataset we are going to explore and discover how certain elements and variables interact with each other. After exploring and cleaning the dataset we have, we are going to answer questions like:

* What are the most extreme game outcomes?
* How does the game affect television viewership?
* How viewership, TV ratings and advert costs evolved over time?
* Who are the most prolific musicians in terms of halftime show performances?

**DATA CLEANING**

For the visualizations, we used the Seaborn library and we used graphs like the Histogram to display the distribution of points scored in the Super Bowl across 10 Super Bowl games. A Scatterplot was also so we could plot the average percentage of USA household with a TV that watched the entirety of the Super Bowl and a Regression graph to see the slope.

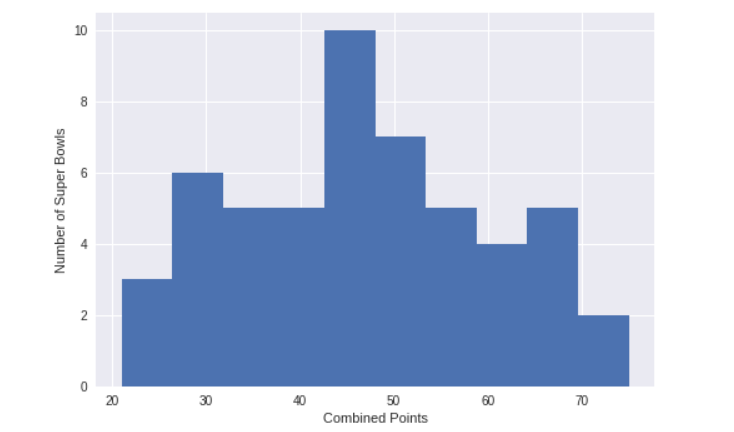
MISSING DATA

Some of the data cleaning methods that were used to clean the data, was that in the halftime musician dataset there were a lot of missing values and that be due to the fact that maybe the TV networks then didn’t really care much about data capturing so we had a list of artists that performed at the Super Bowl but we couldn’t see as to how many songs did they perform on the year they were called to come perform. So, what we did was to filter out certain musicians to address the issue of the missing data

DATA STORIES AND VISUALIZATIONS

## **Combined points distribution**

Here we are looking at combined points for each Super Bowl by visualizing the distribution and we can see the Super Bowl finals with the highest and lowest scores.



## **Point difference distribution**

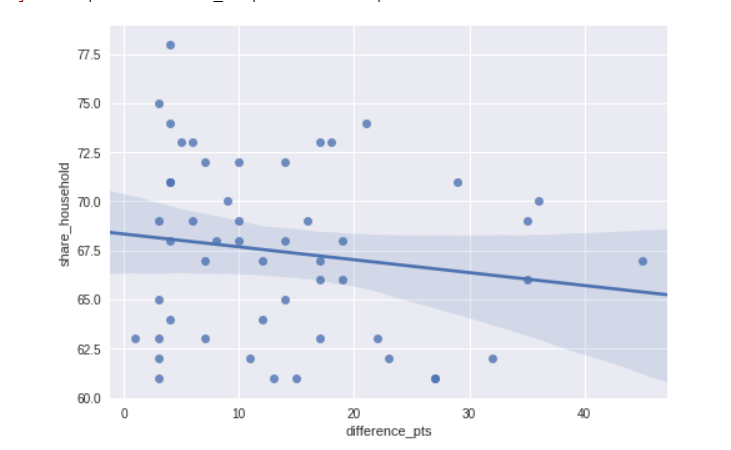
Most combined scores are around 40-50 points, with the extremes being roughly equal distance away in opposite directions. Going up to the highest combined scores at 74 and 75, we find two games featuring dominant quarterback performances.

A picture containing drawing

Description automatically generated

## **Do blowouts translate to lost viewers?**

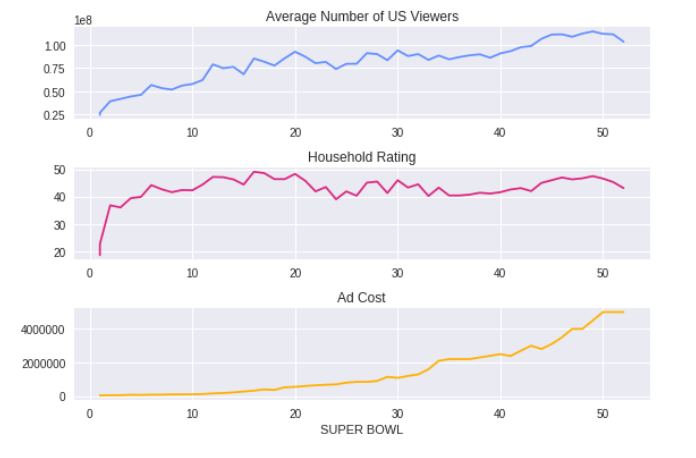
Most of the Super Bowls are close games. Makes sense. Both teams are likely to be deserving if they've made it this far. Let's combine our game data and TV to see if this is a universal phenomenon. Do large point differences translate to lost viewers? We can plot [household share](https://en.wikipedia.org/wiki/Nielsen_ratings) (average percentage of U.S. households with a TV in use that were watching for the entire broadcast) vs. point difference to find out.



## **Viewership and the ad industry over time**

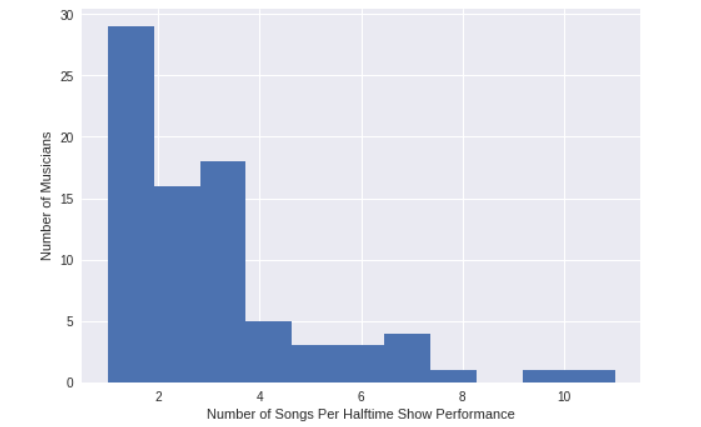
The downward sloping regression line and the 95% confidence interval for that regression *suggest* that bailing on the game if it is a blowout is common. Though it matches our intuition, we must take it with a grain of salt because the linear relationship in the data is weak due to our small sample size of 52 games.

Regardless of the score though, I bet most people stick it out for the halftime show, which is good news for the TV networks and advertisers. A 30-second spot costs a pretty [$5 million](https://www.businessinsider.com/super-bowl-commercials-cost-more-than-eagles-quarterback-earns-2018-1) now, but has it always been that way? And how have number of viewers and household ratings trended alongside ad cost? We can find out using line plots that share a "Super Bowl" x-axis.



## **Halftime shows**

We can see viewers increased before ad costs did. Maybe the networks weren't very data savvy and were slow to react. Another hypothesis: maybe halftime shows weren't that good in the earlier years? The modern spectacle of the Super Bowl has a lot to do with the cultural prestige of big halftime acts. It turns out Michael Jackson's Super Bowl XXVII performance, one of the most watched events in American TV history, was when the NFL realized the value of Super Bowl airtime and decided they needed to sign big name acts from then on out. The halftime shows before MJ indeed weren't that impressive, which we can see by filtering our halftime\_musician data.



**THIS REPORT WAS WRITTEN BY : MASIBONGE MASINGA**

