

## Background

This project is inspired by the [Oakland Athletics](#) baseball team and its [general manager Billy Beane](#). Its focus is the team's analytical, evidence-based, sabermetric approach to assembling a competitive baseball team despite Oakland's small budget.

## Data Set

Data is collected from a database on baseball teams, players and seasons curated by Sean Lahman available at <http://www.seanlahman.com/baseball-archive/statistics/>

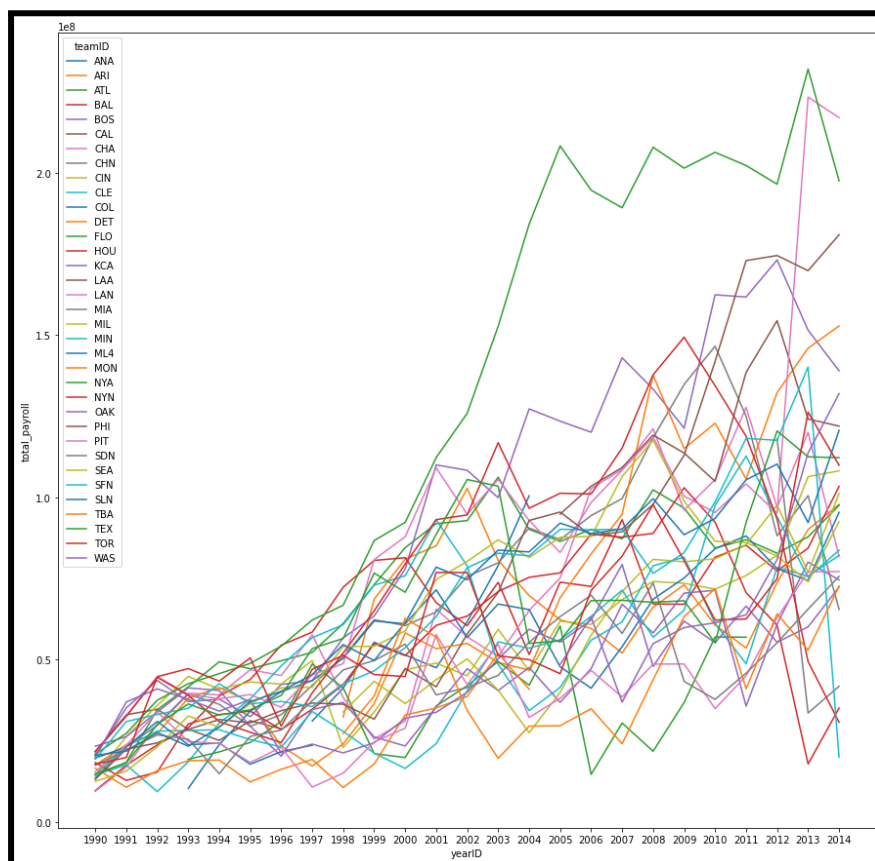
It is a Sqlite database.

## Procedure/Steps

Check the Readme.md file.

## Analysis

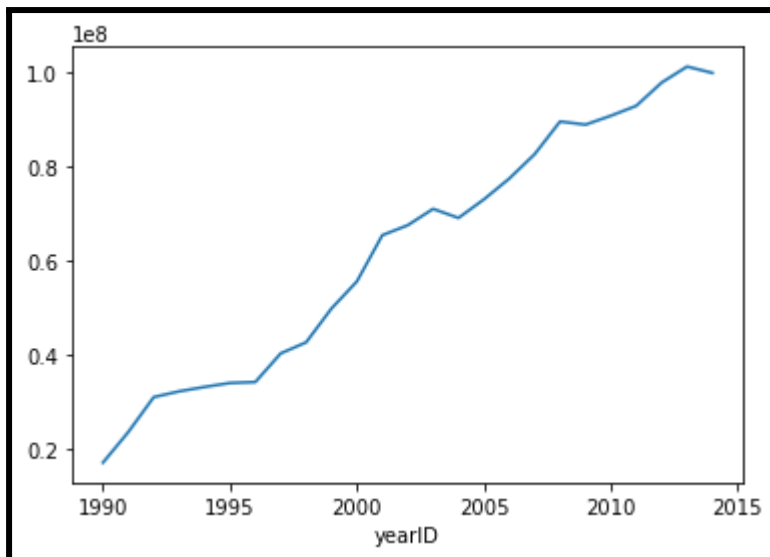
### Plot 1



**Observation:** The plot shows the gradual increase in Total Payroll or budget over the years for all individual teams.

**Analysis:** Most if not all teams have seen an increase in Total Payroll. Interestingly the standard deviation has also increased. Near 1990, mostly all teams were around

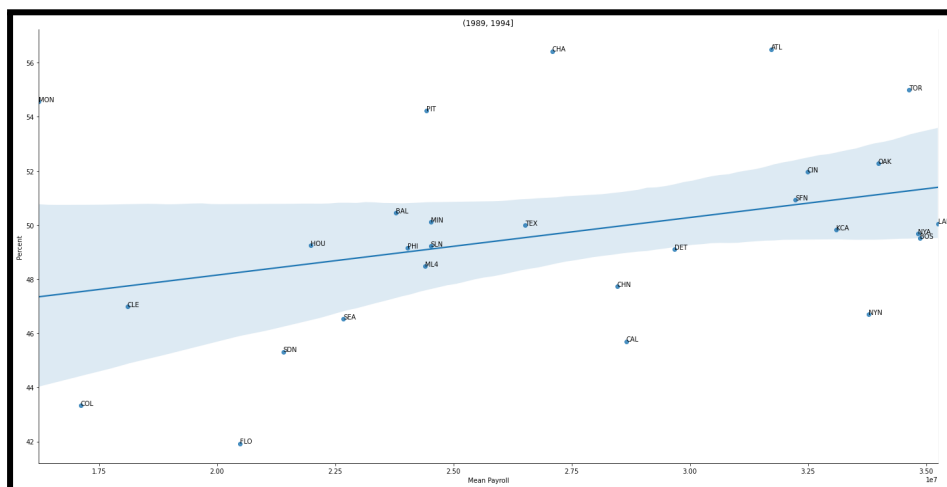
### Plot 2



**Observation:** The plot shows the gradual increase in Total Payroll or budget's mean for all teams over the years .

**Analysis:** In the previous plot we could analyse that there was an increase over time over all. In order to make it clearer I plotted the mean of Payrolls over the years and it proves my previous analysis right.

### Plot 3

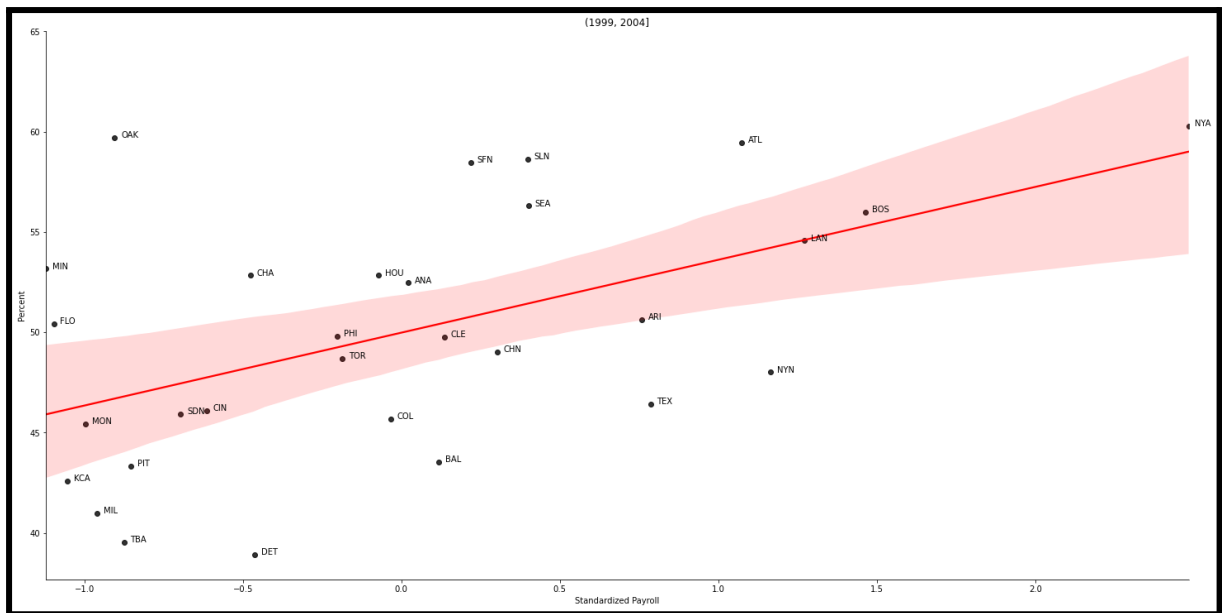


There are five plots for this analysis. I have shown one above (1990-1994). The data frame was sliced into 5 years intervals and the mean payroll was plotted against the mean winning percentage for those 5 years. The teams are data points and there is a regression line showing the common trend.

**Observation:** There is an increasing trend for all year groups. But what is interesting are the outliers, especially with low total payroll.

**Analysis:** Between 1999-2004 we see Oakland posing as a clear outlier with one of the lowest total payroll but the highest winning percentage. This was the golden time for the team but for various reasons (maybe the strategy became public or change in ownership) over the next decade, however, this did not normalize.

**Plot 4**

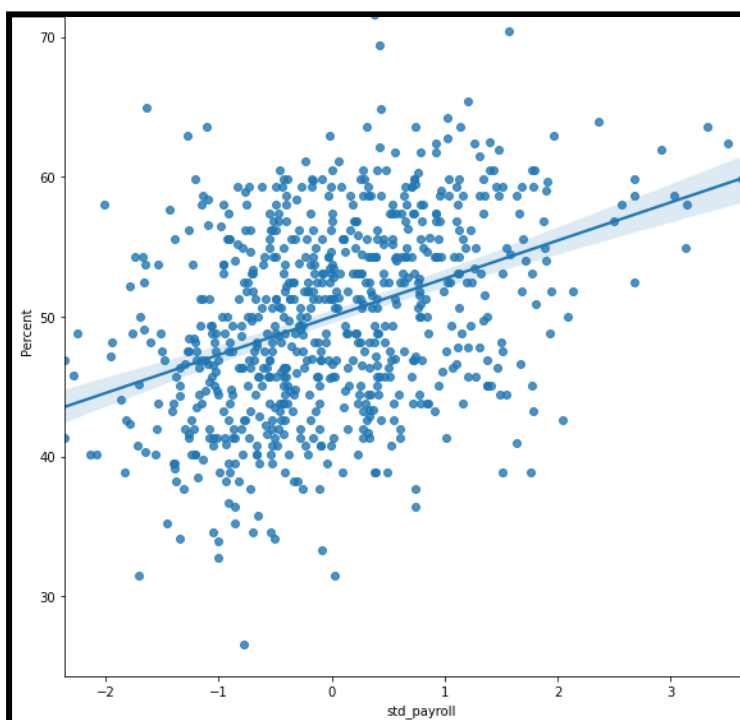


There are five plots for this analysis as well. It is very similar to the previous one. The data frame was sliced into 5 years intervals and the mean standardized payroll was plotted against the mean winning percentage for those 5 years. The teams are data points and there is a regression line showing the common trend.

**Observation:** There is an increasing trend for all year groups. But what is interesting are the outliers, especially with low total payroll.

**Analysis:** Between 1999-2004 we see Oakland posing as a clear outlier with one of the lowest standard total payroll but the highest winning percentage. These plots are still preferable because it plots a standardized payroll variable.

**Plot 5**



**Observation:** The plot shows the positive relation between Standard Payroll and winning percentage.

**Analysis:** We can say that if a team has a greater budget, their chance of winning is greater.

**Plot 6**



**Observation:** There is not an overall observable trend however Oakland shows an efficiency trend over time that matches with our analysis so far.

**Analysis:** From 1997 to 2003-2004 Oakland shows remarkable efficiency in comparison to other important teams. It is indicating the time when the Moneyball stats were being implemented.

## **Conclusion**

My analysis aimed to show the normal trend and relation between attributes of a team. I conclude that there is an increasing relationship between Payroll (total, mean, standardized) with the winning percentage. Money obviously brings in more amenities, players, coaches to create a better team, so it is not surprising.

It is the outliers that were interesting. For this project's purpose, we plotted graphs that showed how in a certain period of time (1997/1999- 2004/2006) Oakland performed extremely well despite their low budget.

**Fun fact:** The team is most famous for winning 20 consecutive games between August 13 and September 4, 2002.