Abdulvahap Yiğit -212970649

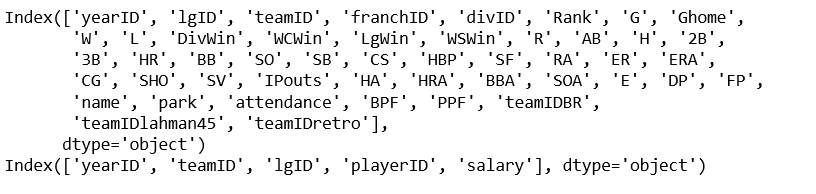
CS 240 Final Project

# Part 1

After all brainstorming with myself I thought it will be good claim “Is there any relationship with total paid salary and win counts in years”. So that’s why I thought that in dataset I have to use teams and salaries.

# Part 2

In part 2 for showing my colums. I wrote simple code script which I learnt in course, and I got that result.

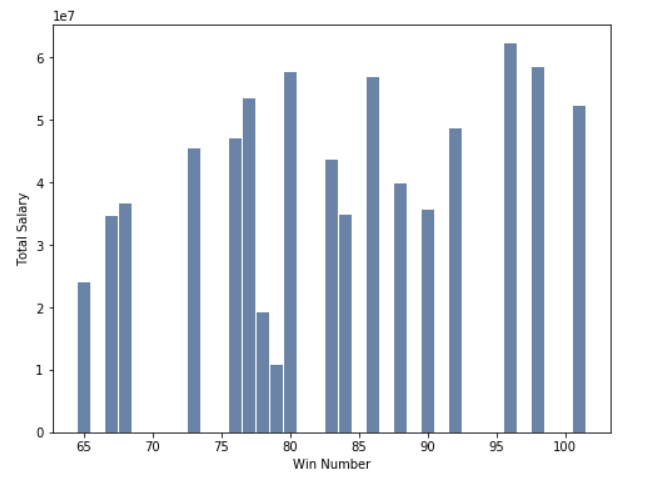


So in above columns I mostly used those columns ; salary, W, teamID, and yearID

# Part 3

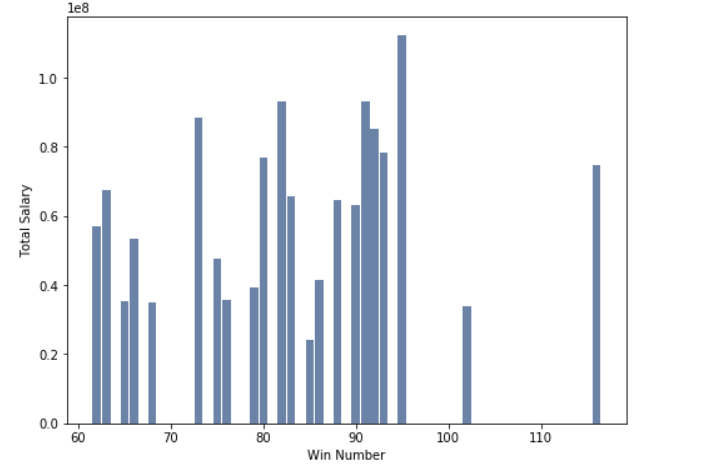
I think that part most important part because I am showing important dataset which as virtualized.

So finally I got those results in my code script ;



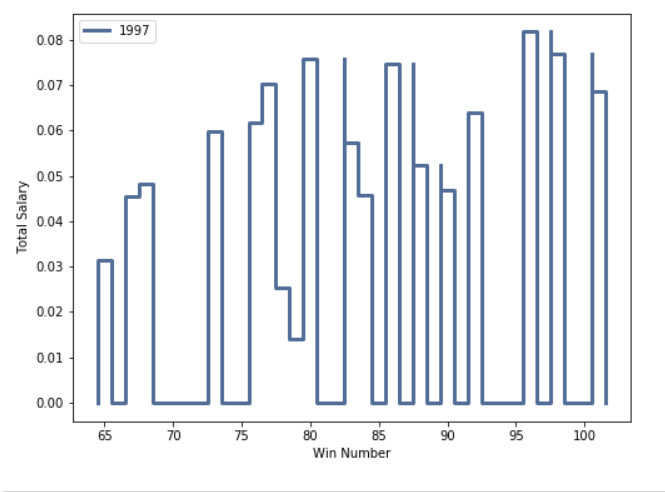
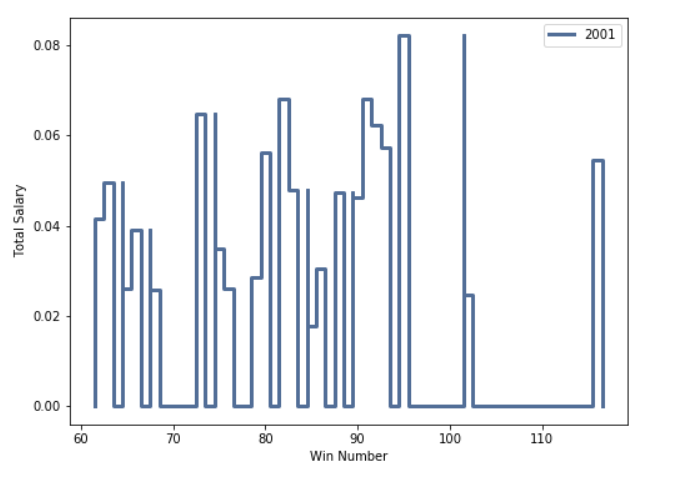
So that’s the histogram which answer that how much team paid to salary and how many win number in 1997.

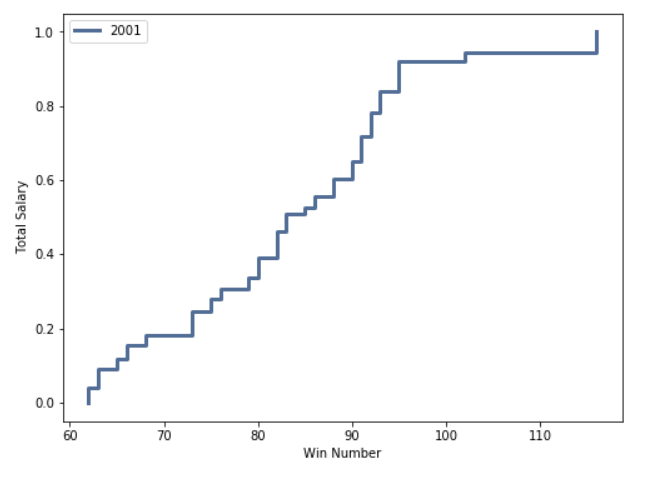
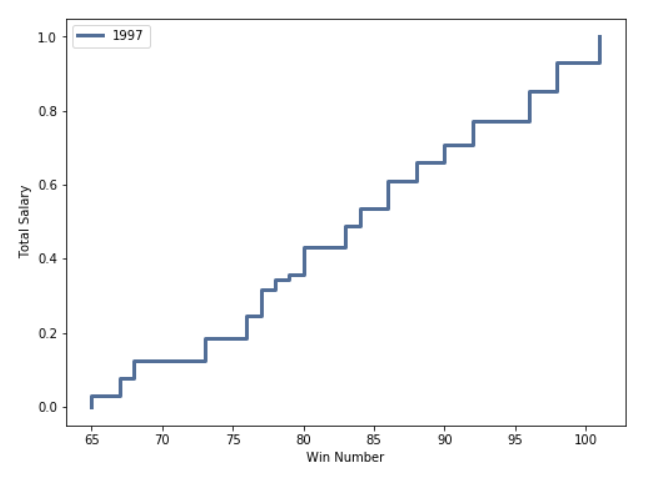
As we can say that team which paid most salary is not the taking most win number.



In 2001 we almost have same previous histogram.

And also we can observe our data in pmf and cdf form.



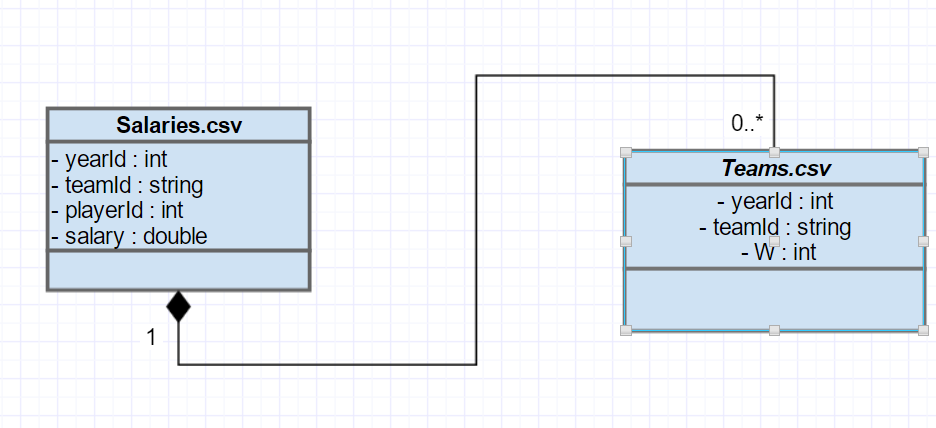


# Part 4

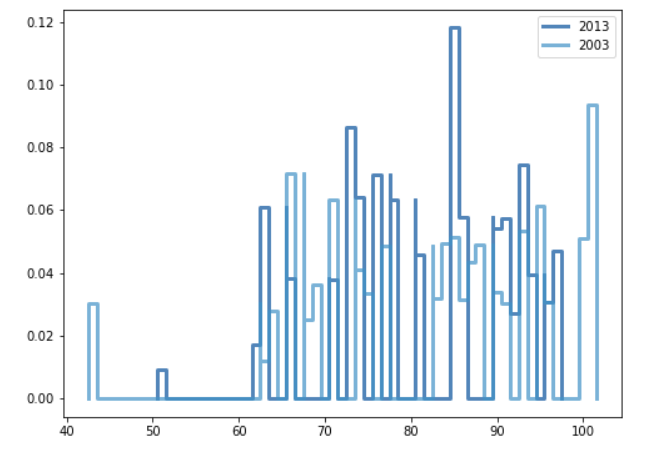
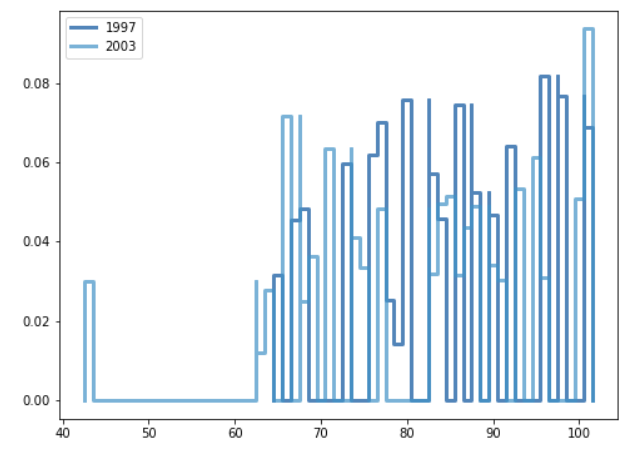
For this project I thought that exponential distributions is the best modelling . Because of my goal is comparing data in some time-interval and proof some differences in year.

# Part 5

I built one relationship between “salaries.csv” and “teams.csv” I pick up salary column in salaries dataset and I pick up team Win counts in teams dataset. So I make this uml for understand better this relationship.



# Part 6



In my jupyter notebook file I explained better but here two visualize that comparing between 1997-2003 and 2003-2013.

# Part 7

In concluison my analysis that for paying more salary to players is not make you are winner. But also we can say that less salary is not making winner. Maybe some others parameters(form,injuries etc.) can effect those result.

End of my analysis I can say that more money does not mean the more win in the baseball game.