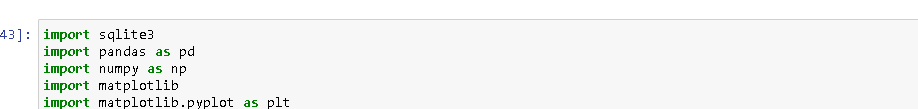
CO

by: omar179771 & omar 12468

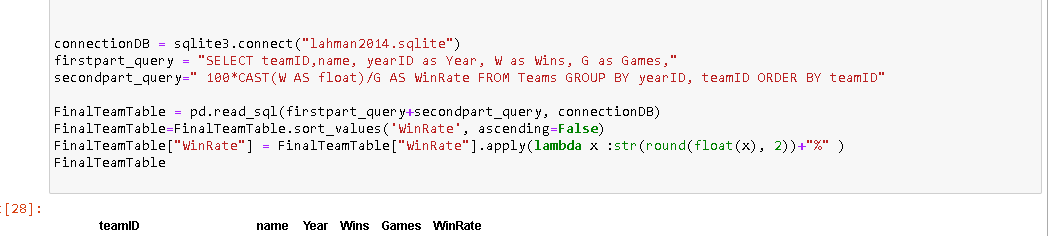
Assignment two

data science



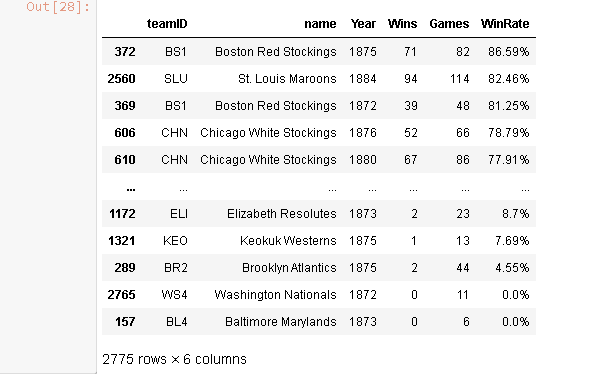
First of all our importation

* Sqlite3 to read our SQLite database file
* Pandas to create tables and make statistics on data
* Numpy used in our cate to create null columns
* Matplotlib .lib and matplotlib pyplot plt used to create plots



Calling DB and win percentage by query

* connectionDB: to connect to the sqlite database by connect function hold parameter sqlite file name and extention and must be in the seam folder of our notebook or it can be another folder but must write the path of the file example connect(“C:/user/username/downloads/{sqlfile}.sqlite”)
* query is select teamid, team name , yearid then I have make 100\* cast”to change type from int to float”( w(wins) as float) as winRate to give here colamn name grouped by teamid and yearid to prevent calculate wins of another team or in the same year
* pandas make wonderful function to call a query and get data in pandas table the parametars query and connection variable
* arrange rows by win rate
* round winrate because win we make the equation it reaches 8 and 9 decimals to I round it to only 2 and convert it to string to boot percentage sine

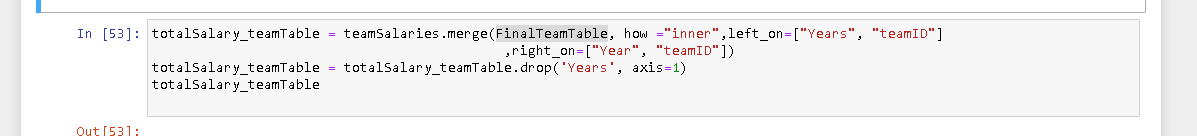




Team salaries for each team in each year

* we get team id , year id , and sum(salary)”sum to get the total of colamns” from table salaries and we also group by year id and team id and order it by teamid alfapatical

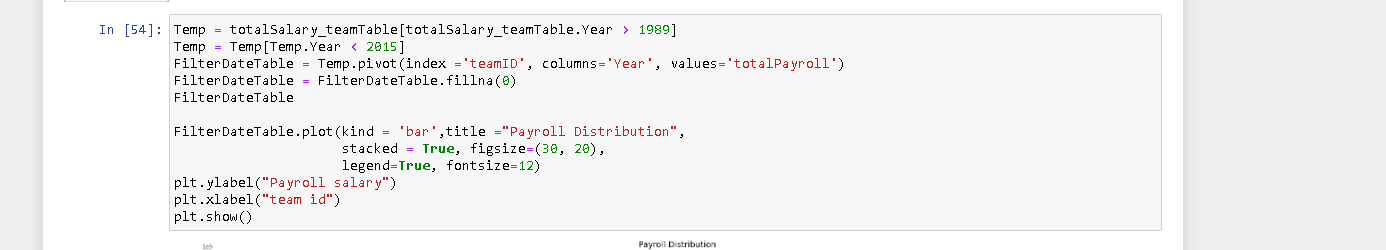




Merge to tables

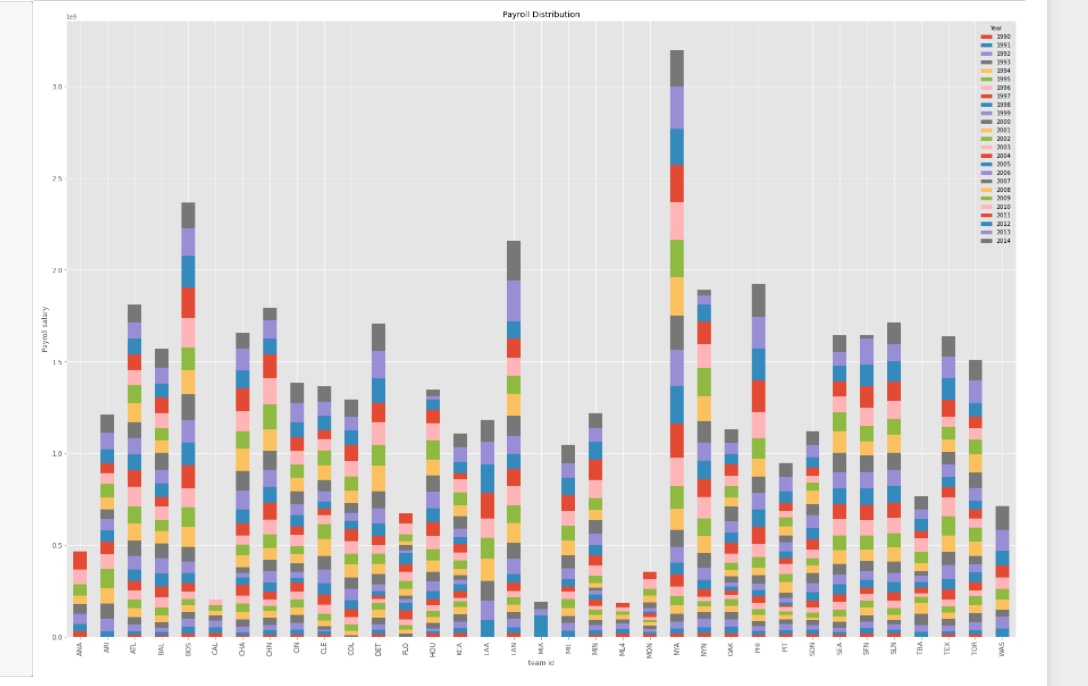
* merge is tool like relations in the database but for pandas table which it sort and merge depended on columns you select In the 2 tables so I have to magre table teamsalary on finalTeamTable because salary have more rows and left on team\_salary and right on finalTeamTable and join type inner

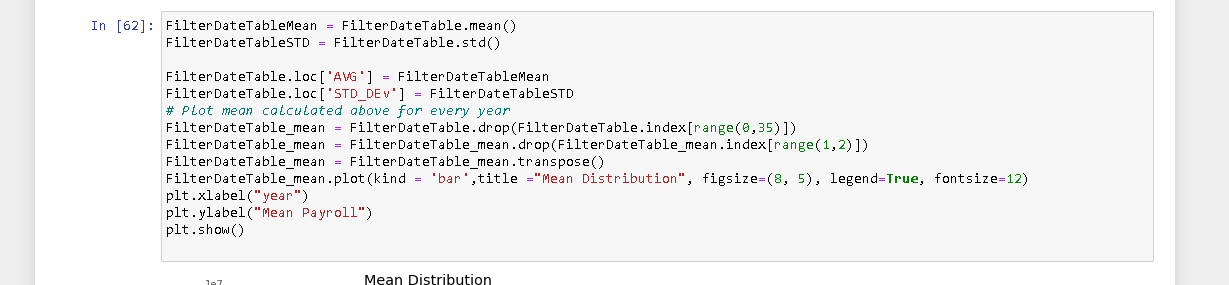




Set date our needed date

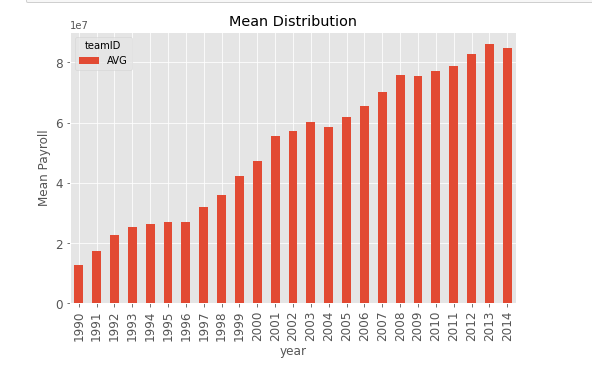
* First to get all rows have column year greater than 1989
* Second get all rows have column year greater than 2015
* Fillna to replace null cells with value I have set it 0
* Draw bar plot

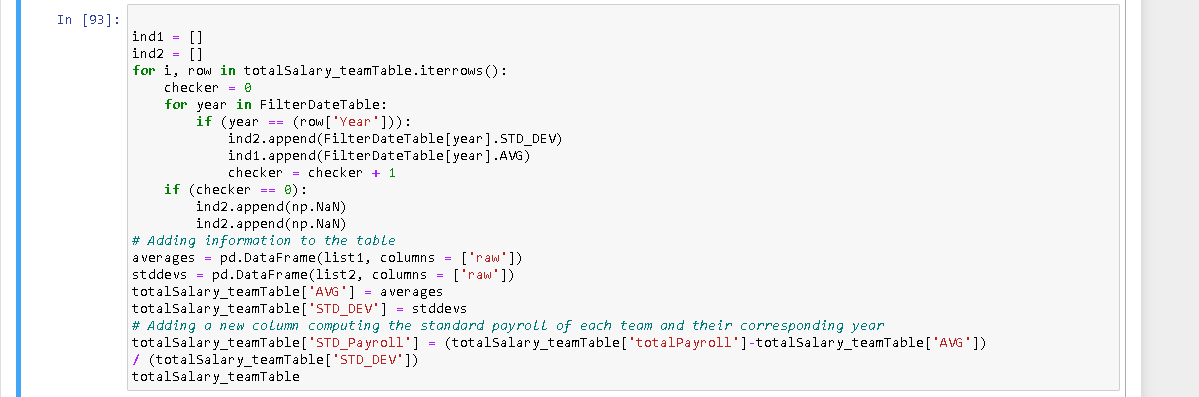




Second plot mean payroll on year

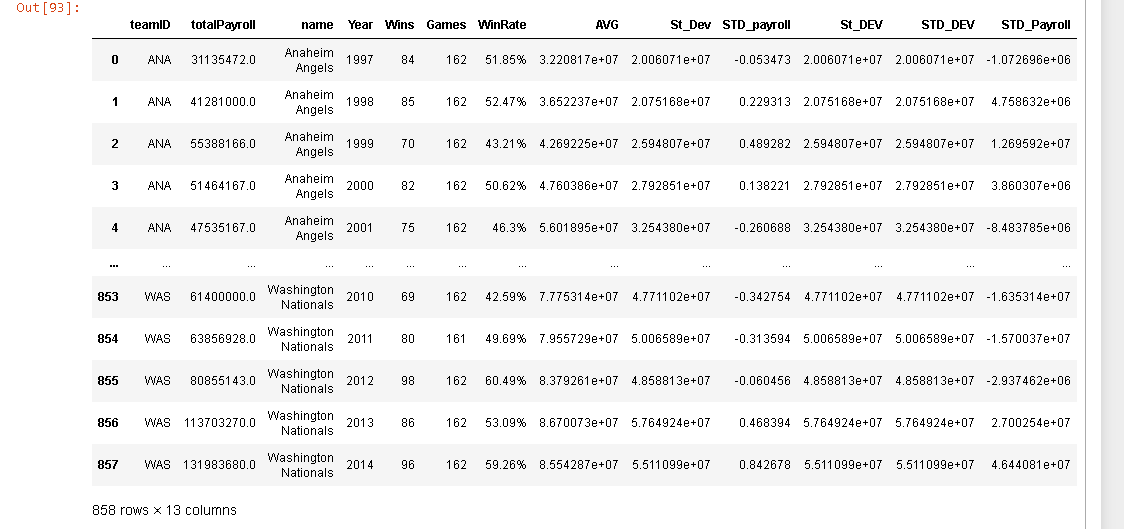
* Get mean and standard deviation
* Set it in columns at the table
* Select index to drop to set only year and payroll
* Then transpose data

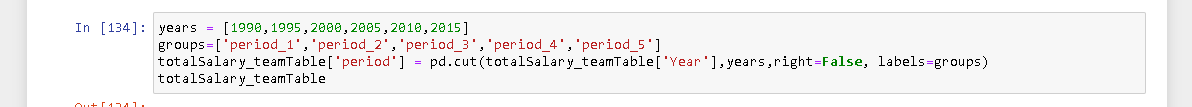




Set avg, std and STD\_payroll in the tables

* It depend on the last plot
* Append in the list by last avg and std we set it before
* Finally get total by totalpayroll / avg / std\_dev





divided it to periods

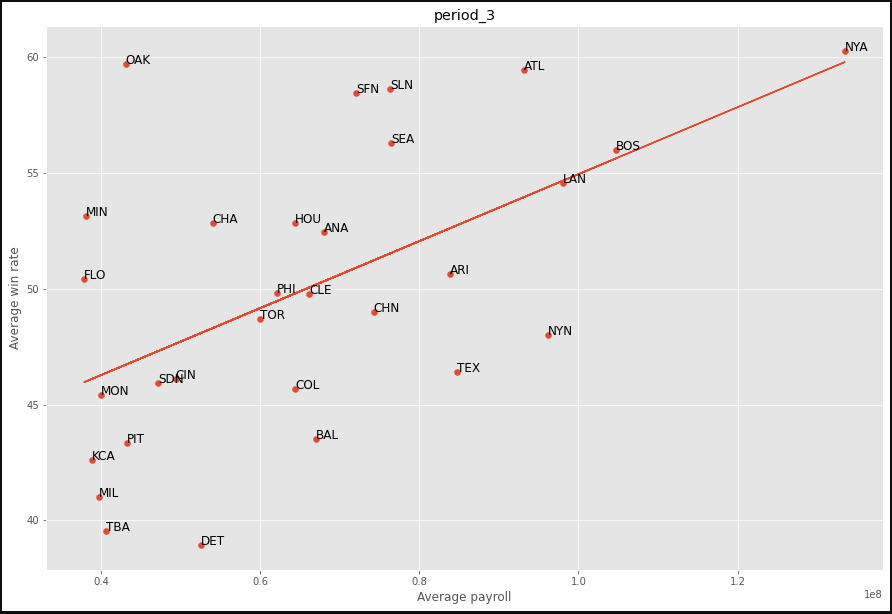
* We used pandas cut function to divide to periods using bins with hold years that bins must be number to devid it. For example period\_1 is 1990<x and 1995>x

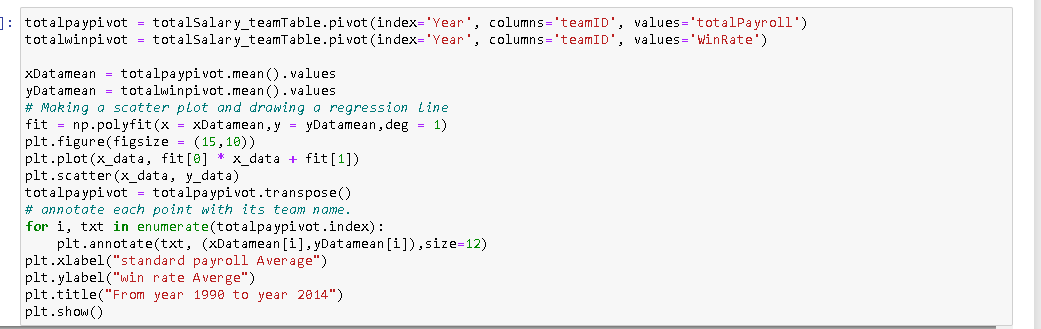




* We get the mean of totalpayroll and winrate depending on year and team id
* Then polyfit the ploy py butting in it the mean values
* Seond for loop to put names on the plot
* This code same as std\_payroll and winrate

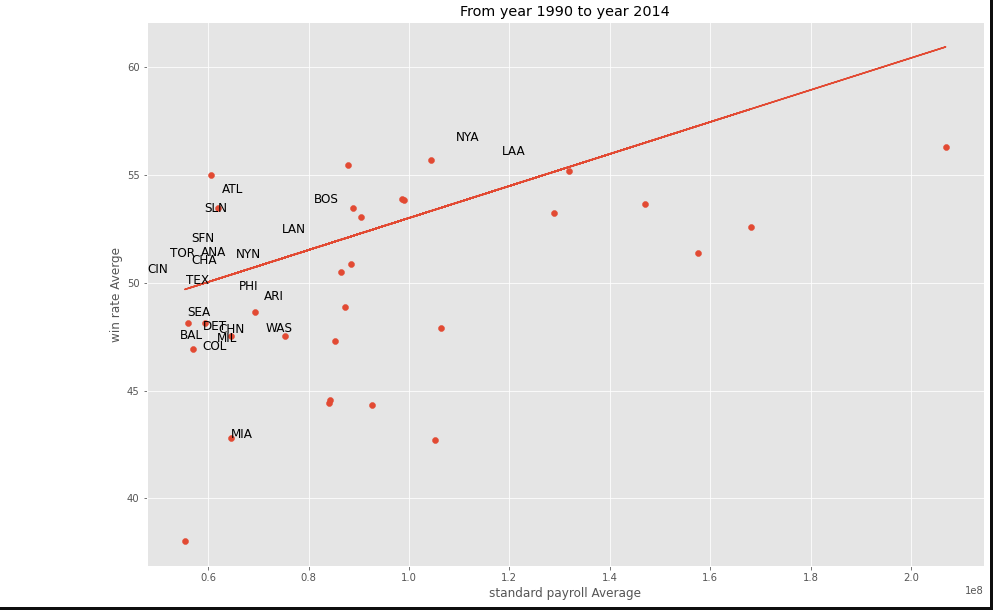
Example from the five periods

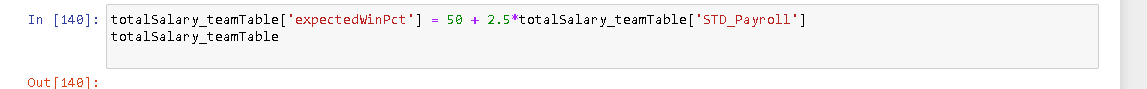




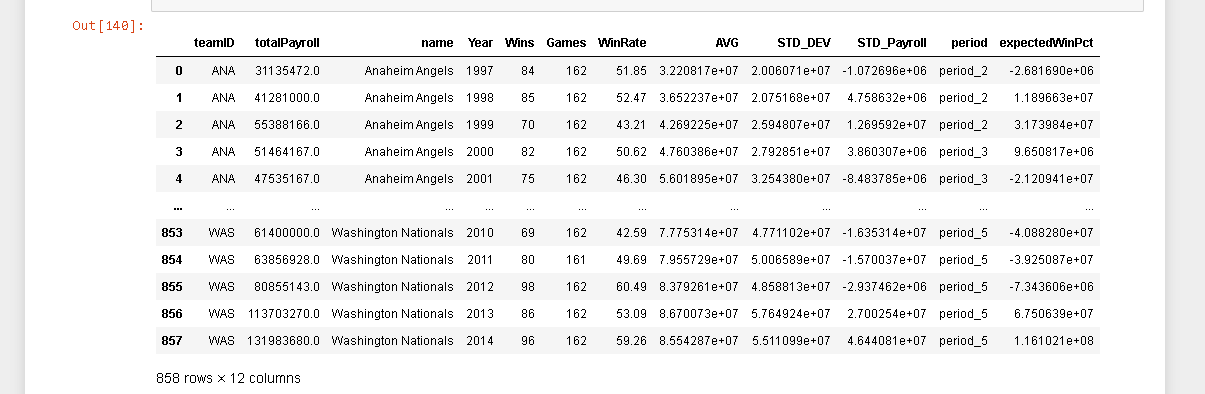
All years plot on winrate and payroll

* Same as periods but without for loop





Get expected win rate





Get team efficacy in plot

* we have make concat with all the team rows

