

Instructions:

- **After completing the assignment, please submit your .ipynb file to NYU Classes with the following naming convention: Lastname_Firstname_NetID_ProblemSet# (ex. Smith_John_js123_ProblemSet2)**
- Submit your answers in a **Jupyter notebook** with **proper markdowns to indicate problem numbers.**
- **Write the questions in markdown before you provide your answers.**
- **When copying the dictionary or any values directly from this file, make sure that all the quotations and brackets are in the right form in Jupyter Notebook. (Especially for string quotations – sometimes if you copy directly from a pdf file, the quotation breaks and it won't show up properly as a string in Jupyter Notebook)**
- See Grading Guidelines under Problem Set 1 Instructions on NYU Classes.
- Before getting into the problems, import `all_data_master.csv`, and replace all `\N` values with `NaN`. Name this data frame as `"all_data"`
- For problems 1 to 6 use `all_data`, so do not change this data frame at any point
- For problems that ask to order by a variable always use ascending order unless stated otherwise
- For problem 6 the overall median is the median of all salaries in `all_data`
- For problem 7 and 8 import csv files `core_data` and `salary_grid` into data frames `employee` and `salary` respectively. From `employee` drop rows where **all** fields are null (Carries credit)
- In-line comments are preferred for this assignment but not mandatory
- No explanations are expected at the end of answers, unless requested

Problems:

1. Display total number of job postings in each year. Print the year that had most jobs. Plot a simple line graph to see if jobs rise with each passing year.
2. Display mean salary per year for the company Wells Fargo in a single data frame (company, year, mean_salary). Plot a graph to determine whether Wells Fargo mean salaries are on the rise with every passing year.
3. Display standard deviation in salaries for the states AZ, TX and DC in descending order. Now visualize this data in a bar chart.
4. Display `all_data` without those states that have less than 1000 job postings. Final data frame must include all columns as the original data frame.
5. For each state, find the company that posted the job with highest salary (among all job postings in that state alone). Final data frame must have columns `job_id`, `company`, `salary`. There will be only one record per state.
6. Display `all_data` without those companies whose highest salary was lower than the overall median. Final data frame must include all columns of the original data frame.
7. Get salary information for all employees. Display the employee name, state, age, position and `Hourly_Max` salary offered.
8. Who are the top 20 highest paid employees based on the `Hourly_Max` salary column? Print the percentage of top 20 employees that fully meet their performance score.