

U.S. SOFTWARE SALARY ANALYSIS



INTRODUCTION

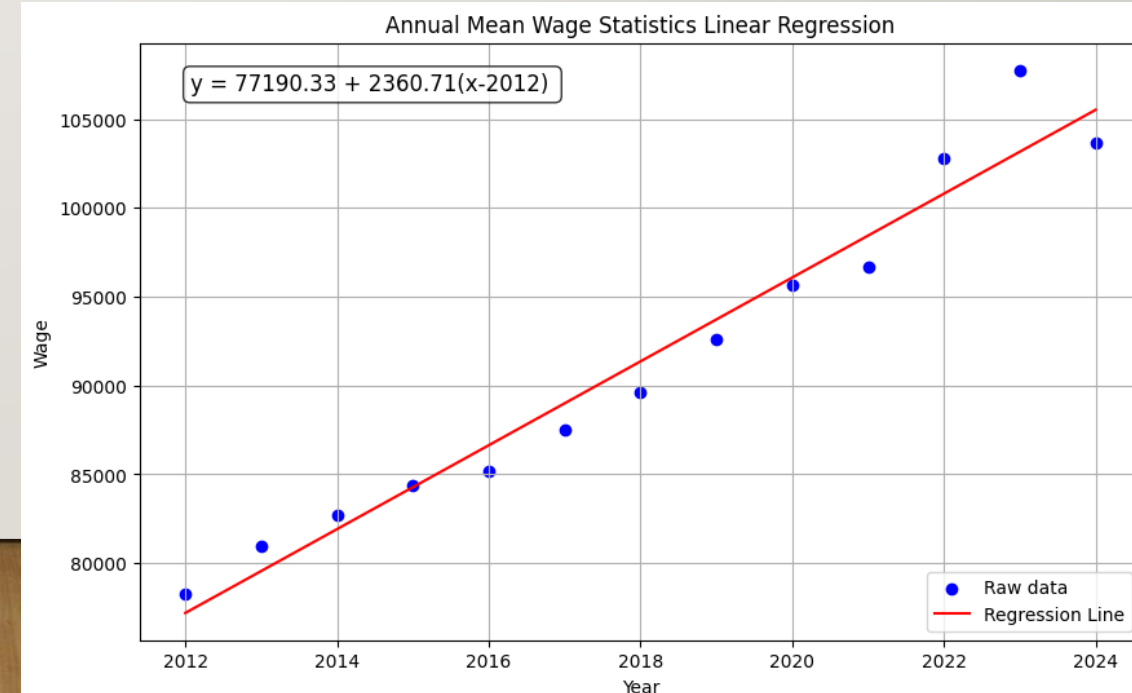
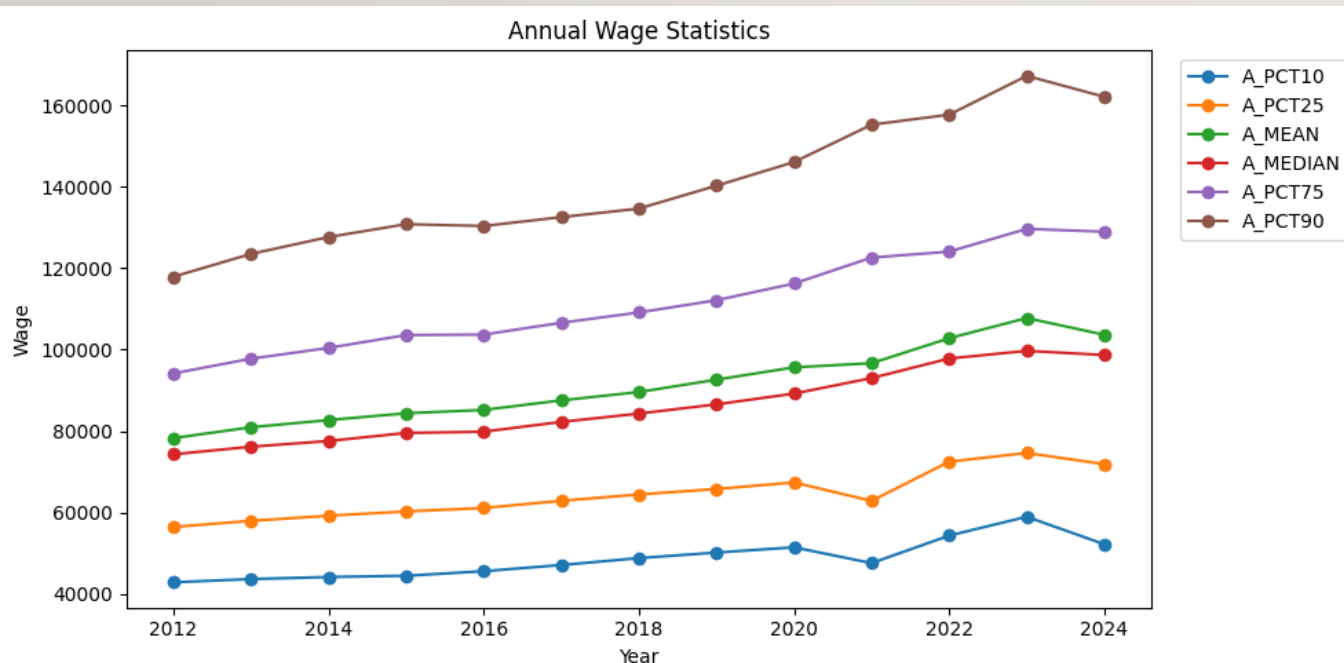
- This project analyzes U.S. programmer salary statistics to support IT career planning.
- I analyze the relationships between salary and time, city, local purchasing power, etc., using data from Kaggle, levels.fyi and U.S. bureau of labor statistics, using multiple data collecting and preprocessing skills.

DATA SOURCE AND APPROACH

Data Source#	Name / Short description	Source URL	Type	List of fields		Format
1	Annual Wage Statistics	https://www.bls.gov/oes/tables.htm	file	1. Year 2. A_PCT10 3. A_PCT25 4. A_MEAN	5. A_MEDIAN 6. A_PCT75 7. A_PCE90	CSV
2	U.S. Software Developer Salaries	https://www.kaggle.com/datasets/thedevastator/u-s-software-developer-salaries	API	1. city 2. Salary	3. num_jobs 4. purchasing_power	CSV
3	Software Engineer Salary	https://www.levels.fyi/locations?jobFamily=Software+Engineer	Web page	1. city 2. median	3. min 4. max	CSV

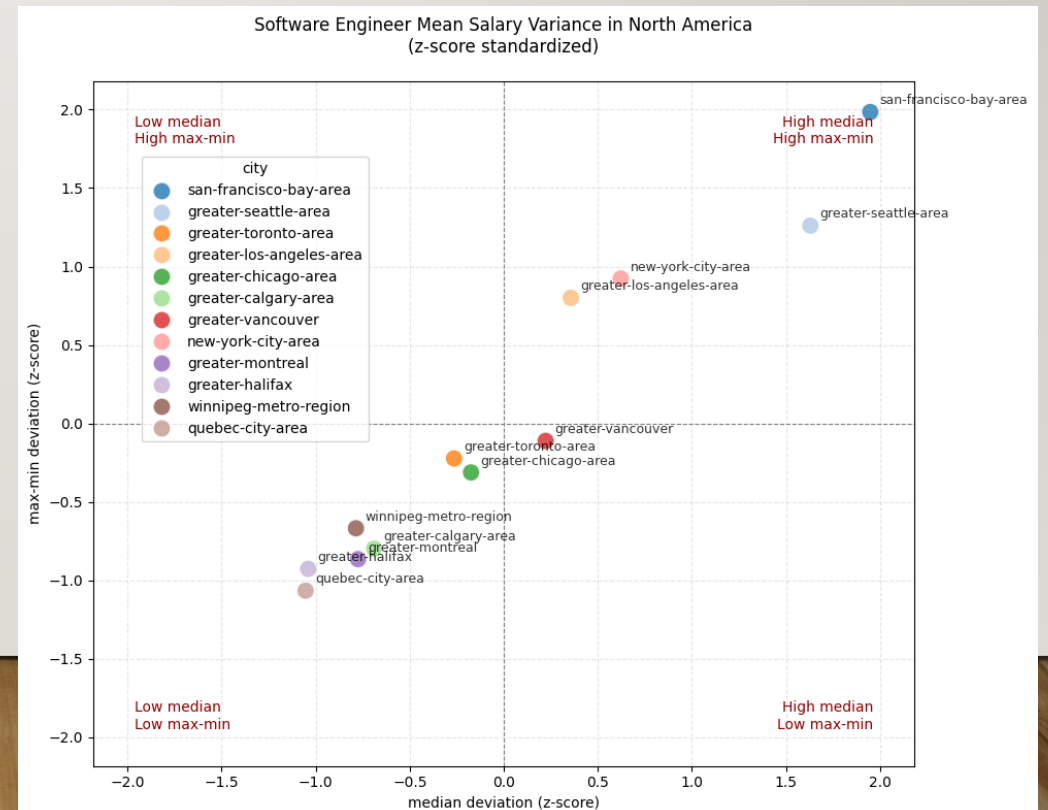
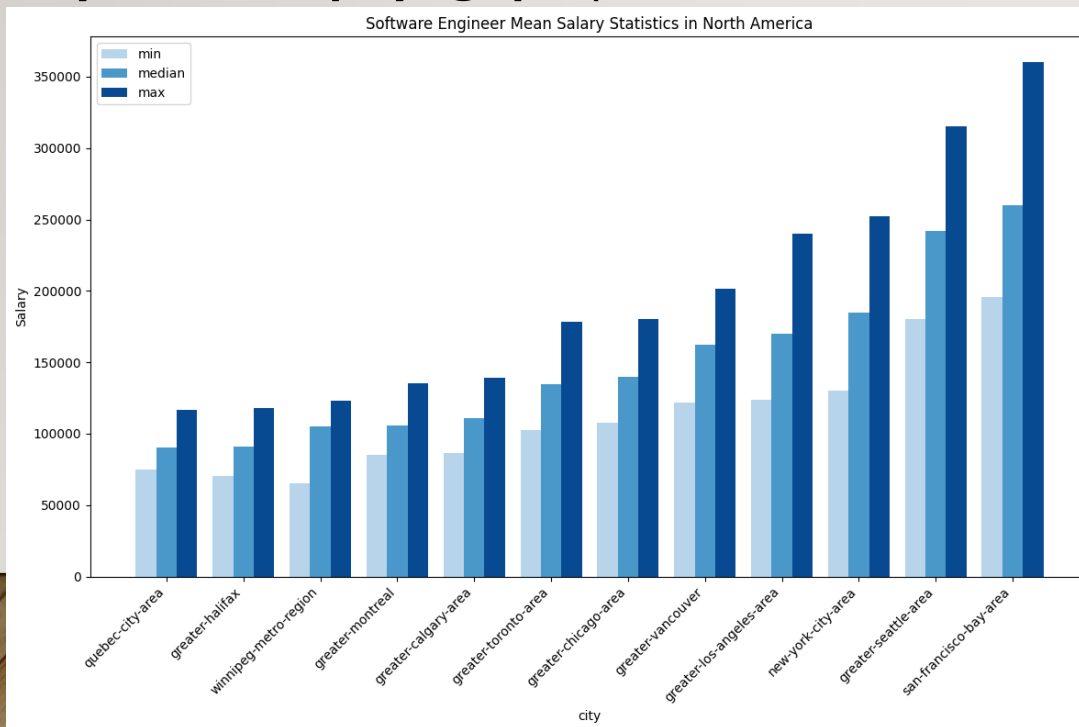
SUMMARY OF THE RESULTS — TREND OF ANNUAL WAGES

1. The salary is increasing overall in different indicators (e.g. percentiles, median).
2. Salary is proportional to years and can be fitted with linear regression, with OLS analysis showing high accuracy in the fit.



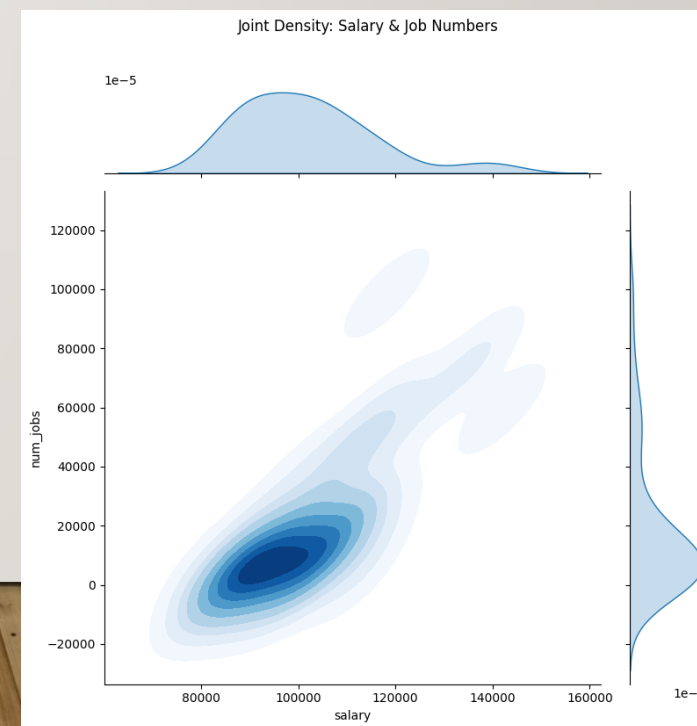
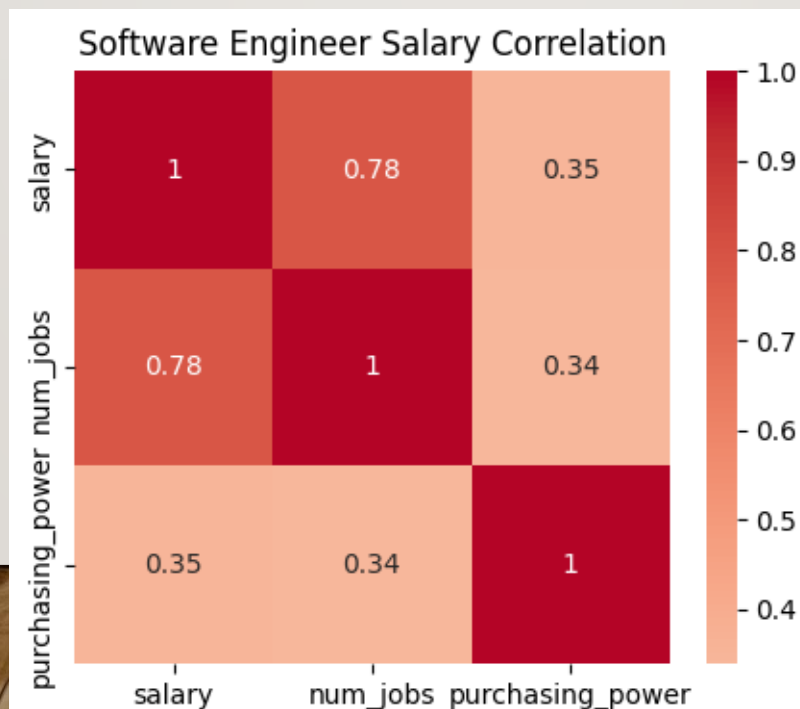
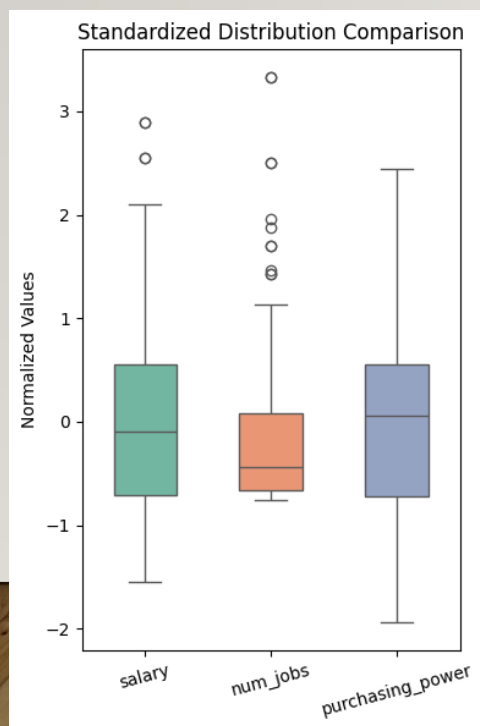
SUMMARY OF THE RESULTS — SALARY DISTRIBUTION BETWEEN CITIES

1. Salaries vary significantly across cities.
2. Higher median salary correlates with greater pay disparity, i.e., higher median salary, wider pay gap. (Pearson's $r = 0.97$)



SUMMARY OF THE RESULTS — SALARY, NUM_JOBS AND PURCHASING_POWER

1. Job opportunities vary widely by city, but salaries and purchasing power differ less.
2. Salary correlates with job numbers ($r=0.78$), with weaker ties to other factors, evident in the KDE plot's upward-tilted elliptical pattern.



DATA COMBINING AND JOINT ANALYSIS — SALARY, YEAR, CITY

1. Merge data source 2 (2022) and source 3 (2025) with same cities for data combining, find annual changes in different cities.
2. Compare with enhanced data (added 2025) source 1 by linear regression prediction, check validity of the data sources.

