# U.S. Software Developer Salaries

#### **1. Dataset Selection:**

**About this dataset**

This dataset provides an extensive look into the financial health of software developers in major cities and metropolitan areas around the United States. We explore disparities between states and cities in terms of mean software developer salaries, median home prices, cost of living avgs, rent avgs, cost of living plus rent avgs and local purchasing power averages. Through this data set we can gain insights on how to better understand which areas are more financially viable than others when seeking employment within the software development field. Our data allow us to uncover patterns among certain geographic locations in order to identify other compelling financial opportunities that software developers may benefit from

**How to use the dataset**

This dataset contains valuable information about software developer salaries across states and cities in the United States. It is important for recruiters and professionals alike to understand what kind of compensation software developers are likely to receive, as it may be beneficial when considering job opportunities or applying for a promotion. This guide will provide an overview of what you can learn from this dataset.

The data is organized by metropolitan areas, which encompass multiple cities within the same geographical region (e.g., “New York-Northern New Jersey” covers both New York City and Newark). From there, each metro can be broken down further into a number of different factors that may affect software developer salaries in the area:

* Mean Software Developer Salary (adjusted): The average salary of software developers in that particular metro area after accounting for cost of living differences within the region.
* Mean Software Developer Salary (unadjusted): The average salary of software developers in that particular metro area before adjusting for cost-of-living discrepancies between locales.
* Number of Software Developer Jobs: This column lists how many total jobs are available to software developers in this particular metropolitan area.
* Median Home Price: A metric which shows median value of all homes currently on the market within this partcular city or state. It helps gauge how expensive housing costs might be to potential residents who already have an idea about their income/salary range expectations when considering a move/relocation into another location or potentially looking at mortgage/rental options etc.. 5) Cost Of Living Avg: A metric designed to measure affordability using local prices paid on common consumer goods like food , transportation , health care , housing & other services etc.. Also prominent here along with rent avg ,cost od living plus rent avg helping compare relative cost structures between different locations while assessing potential remunerations & risk associated with them . 6)Local Purchasing Power Avg : A measure reflecting expected difference in discretionary spending ability among households regardless their income level upon relocation due to price discrepancies across locations allows individual assessment critical during job search particularly regarding relocation as well as comparison based decision making across prospective candidates during any hiring process . 7 ) Rent Avg : Average rental costs for homes / apartments dealbreakers even among prime job prospects particularly medium income earners.(basis family size & other constraints ) 8 ) Cost Of Living Plus Rent Avg : Used here as one sized fits perspective towards measuring overall cost structure including items

**Research Ideas**

* Comparing salaries of software developers in different cities to determine which city provides the best compensation package.
* Estimating the cost of relocating to a new city by looking at average costs such as rent and cost of living.
* Predicting job growth for software developers by analyzing factors like local purchasing power, median home price and number of jobs available

**Acknowledgements**

If you use this dataset in your research, please credit the original authors.  
[Data Source](https://zenodo.org/record/7412091#.Y9Y3ZNJBwUE)

**License**

**License:**[**CC0 1.0 Universal (CC0 1.0) - Public Domain Dedication**](https://creativecommons.org/publicdomain/zero/1.0/)  
No Copyright - You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission. [See Other Information](https://creativecommons.org/publicdomain/zero/1.0/).

**Columns**

**File: SofwareDeveloperIncomeExpensesperUSACity.csv**

| **Column name** | **Description** |
| --- | --- |
| **Metro** | The metropolitan area of the city. (String) |
| **Mean Software Developer Salary (adjusted)** | The average salary for software developers adjusted for cost of living differences between cities. (Number) |
| **Mean Software Developer Salary (unadjusted)** | The average salary for software developers without adjusting for cost of living differences between cities. (Number) |
| **Mean Unadjusted Salary (all occupations)** | The average salary for all occupations without adjusting for cost of living differences between cities. (Number) |
| **Number of Software Developer Jobs** | The number of software developer jobs in the city. (Number) |
| **Median Home Price** | The median home price in the city. (Number) |
| **City** | The name of the city. (String) |
| **Cost of Living avg** | The average cost of living in the city. (Number) |
| **Rent avg** | The average rent in the city. (Number) |
| **Cost of Living Plus Rent avg** | The average cost of living plus rent in the city. (Number) |
| **Local Purchasing Power avg** | The average local purchasing power in the city. (Number) |

2. Dataset Inspection:

| **Column name** | **Description** |
| --- | --- |
| **Metro** | The metropolitan area of the city. (String) |
| **Mean Software Developer Salary (adjusted)** | The average salary for software developers adjusted for cost of living differences between cities. (Number) |
| **Mean Software Developer Salary (unadjusted)** | The average salary for software developers without adjusting for cost of living differences between cities. (Number) |
| **Mean Unadjusted Salary (all occupations)** | The average salary for all occupations without adjusting for cost of living differences between cities. (Number) |
| **Number of Software Developer Jobs** | The number of software developer jobs in the city. (Number) |
| **Median Home Price** | The median home price in the city. (Number) |
| **City** | The name of the city. (String) |
| **Cost of Living avg** | The average cost of living in the city. (Number) |
| **Rent avg** | The average rent in the city. (Number) |
| **Cost of Living Plus Rent avg** | The average cost of living plus rent in the city. (Number) |
| **Local Purchasing Power avg** | The average local purchasing power in the city. (Number) |

3. Data Visualization:

4. Data Cleaning and Preprocessing:

Address the issues identified during dataset inspection. This might involve handling missing values, dealing with outliers, standardizing formats, and resolving inconsistencies. Document the steps you take and explain your reasoning.

1). Remove irrelevant data

2). Deduplicate data

3). Fix structural errors

4). Deal with missing value

5). Filter out data outliers