

## Academic Honesty

Name(s) of all authors: Guochen

Assignment name: Data Description

Assignment due date: Sep 18

Written/online sources used:

<https://www.ibm.com/docs/en/spss-modeler/18.2.2?topic=data-writing-description-report>

Help obtained (Acknowledgments): None

Add the statement: We confirm that the above list of sources is complete AND that we have not talked to anyone else about the solution to this problem.

## Data Quantity

The dataset, encapsulating information on software developer salaries, is formatted in a CSV file. This dataset is the result of the processing and aggregation of various datasets, each providing different perspectives and insights. The database is substantial, comprising 12 columns that represent various facets of income and expenditure and 77 rows, each detailing the data for a specific city.

## Data Quality

The dataset is characterized by a high degree of relevance to the business questions at hand, including critical attributes such as purchasing power, rent, and house prices, which are pivotal in analyzing the economic landscape for software developers. The dataset is composed of two primary data types: symbolic and numeric. The symbolic data represents city names and their respective states, providing a geographical context to the analysis. The numeric data indicates critical financial metrics such as salaries and prices, offering quantitative insight to evaluate the industry.

## Statistical Insights

Upon computing basic statistics for key attributes, I noticed that a significant insight emerged regarding the rent-to-salary ratio across different career paths. It was observed that the rent constitutes approximately 21.4% of the total salary for a software developer, which contrasts with the 39.5% observed for average salaries across all careers. It underscored a notable discrepancy in the salary structures between software engineers and other professions, highlighting the financial advantages associated with a career in software development. While the current analysis provides a substantial foundation, collaboration with business analysts could potentially unveil deeper insights and nuances associated with this discrepancy.