# STEM OCCUPATION WAGE ANALYTICS

## COMPREHENSIVE REVIEW OF TRENDS ACROSS THE UNITED STATES

SUSHANT THAPA MAGAR
MSC. BUSINESS ANALYTICS

# INTRODUCTION

In an era dominated by rapid technological growth and evolving economic scenarios, distinguishing wage dynamics between STEM and NON-STEM occupations across Metropolitan Statistical Areas (MSAs) presents a unique analytical challenge. This report begins with a holistic view of these disparities, leveraging a robust dataset from the Bureau of Labor Statistics to compare and contrast wages in these divergent sectors. Initial findings highlight not only significant regional wage discrepancies but also the differential impacts of economic factors such as GDP growth and inflation on these sectors.

Building upon these foundational insights, the report narrows its focus to a detailed exploration of STEM occupations. Employing advanced analytical techniques in Alteryx, the study transitions from broad comparative analysis to an intensive examination of the variables influencing STEM wages specifically. The scope of this analysis is comprehensive, designed to unravel the complexities of wage variations, predict future employment trends, and formulate strategies that could influence policy decisions.

This investigative journey is structured to provide a layered understanding of the economic and occupational landscape affecting STEM professionals. It begins with a discussion of the data sources and methodologies employed, followed by a presentation of the findings from the descriptive analysis. Subsequent sections delve into predictive modeling techniques and their outcomes , culminating in strategic recommendations that leverage these insights to propose actionable measures.

The assumptions framing this research—such as the anticipated uniformity of economic impact across different regions and occupations—shape the analytical models and strategies developed. These assumptions, along with a detailed record of the project's evolution captured in the Development Diary, ensure that the analysis remains grounded and reflective of real-world complexities.

## TRENDS AND THEORETICAL PERSPECTIVES

Through a blend of descriptive, diagnostic, predictive, and prescriptive analytics, this report aims not just to illuminate current conditions but also to forecast and influence future trends, providing valuable insights for policymakers, educational institutions, and industry stakeholders dedicated to fostering a resilient STEM workforce aligned with dynamic economic conditions.

The STEM labor market is experiencing a significant surge, with projections from the U.S. Bureau of Labor Statistics indicating a growth rate of 10.8% from 2022 to 2032, surpassing overall employment growth. This increase is driven by the rapid expansion of the tech industry and transformative technologies, alongside supportive policies like the CHIPS and Science Act which promote STEM research and education. Data from the Occupational Employment and Wage Statistics (OEWS) confirms these trends, highlighting key MSAs as centers of STEM activity and growth.

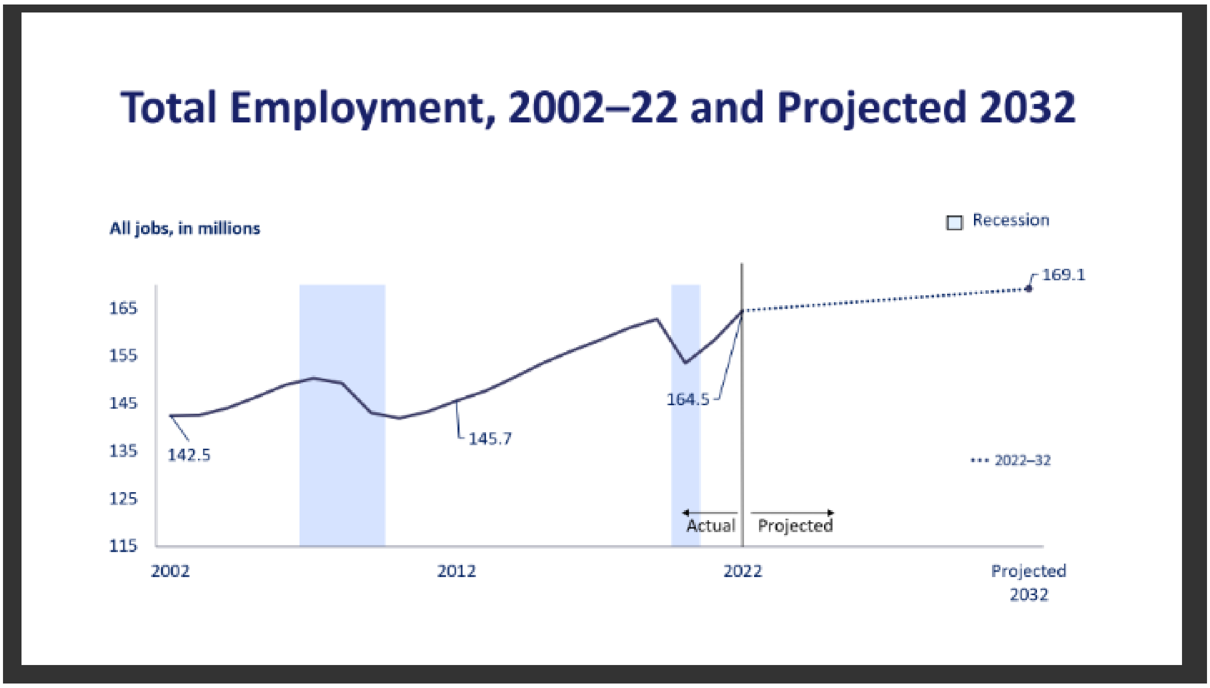



Image Source: BLS

The rise in STEM demand aligns with Human Capital Theory, which suggests that investments in education enhance productivity and earnings, particularly visible in STEM fields where advanced degrees yield higher wages. Labor Market Segmentation Theory also applies, illustrating how specialized skills and regional policies create distinct employment segments within STEM, affecting wage structures and employment terms across different regions.

Traditional economic models struggle to keep pace with the rapid changes in STEM fields, often failing to predict the shifts in job roles and skills demand driven by technological advancements. For instance, the significant increase in jobs requiring high digital skills from 5% to 23% between 2002 and 2016 illustrates these dynamics (Brookings Institution). These models fall short in capturing the nuanced variations in digital skill demands and the broader economic impacts on the STEM labor market.

The initial assumption that detailed SOC data would yield the most precise insights was reevaluated through the research. While useful, this data often missed broader economic impacts like GDP growth and inflation effects on wages, suggesting a need for a more aggregated approach. This insight led to a methodological shift towards incorporating cross-industry data, enriching the understanding and relevance of the findings in depicting the complex nature of STEM employment.

## ANALYTICAL APPROACH

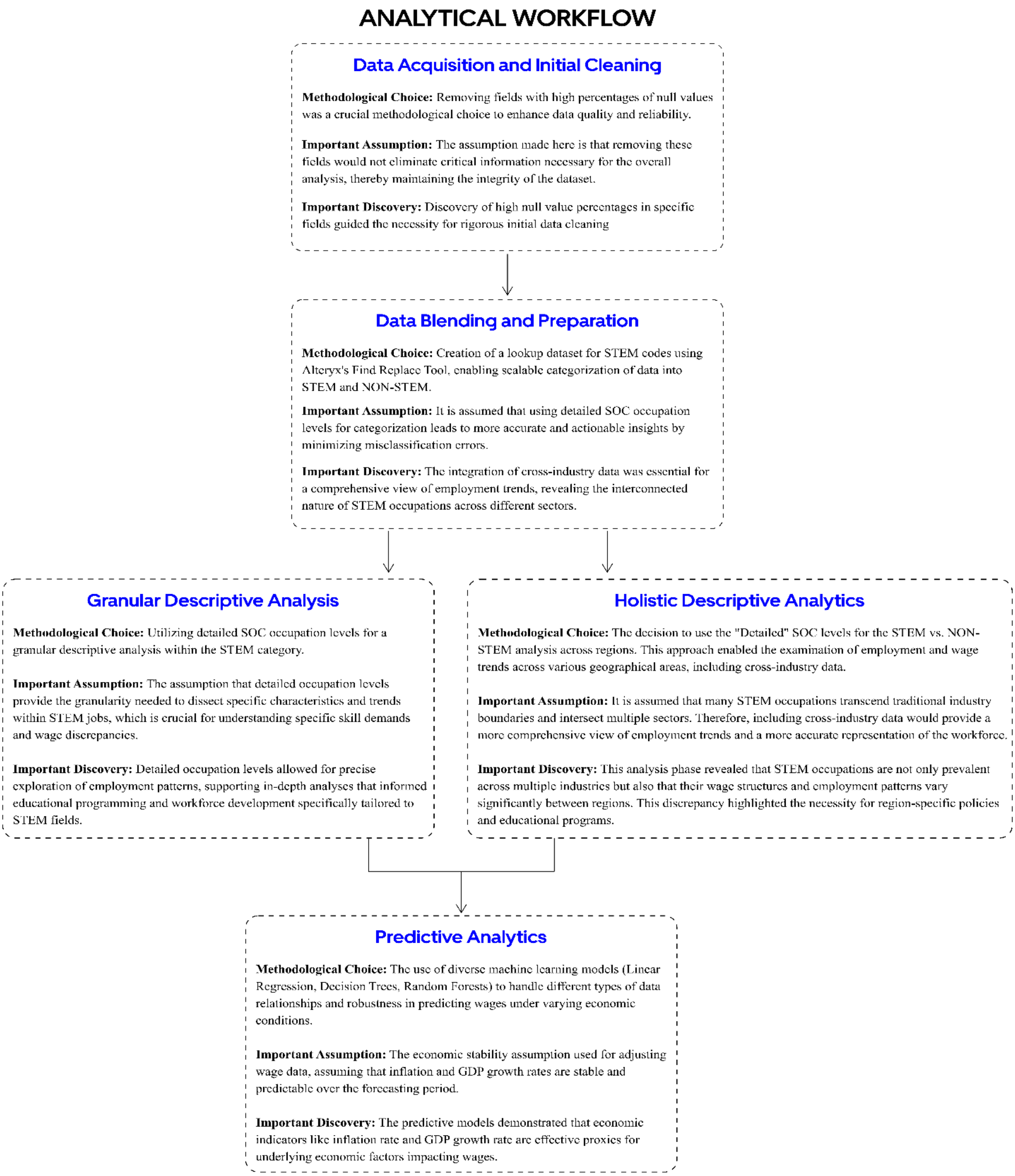
The analytical framework began with thorough data integration using Alteryx, prioritizing data cleaning and preparation to address significant missing values in crucial fields. A lookup dataset for STEM codes facilitated scalable categorization into STEM or NON-STEM, ensuring precise occupation identification. Descriptive analytics provided insights into wage distribution within major Metropolitan Statistical Areas (MSAs), followed by predictive analytics employing robust machine learning models like Linear Regression, Decision Trees, and Random Forests to predict wage differentials effectively.

Ethically, high standards were maintained in data privacy and accuracy, with sensitive information anonymized and adherence to data protection regulations. Transparency was upheld for reproducibility and peer review.

The analysis revealed the substantial economic value of STEM jobs, consistently showcasing higher wages nationwide. Regional disparities in both employment and wage levels underscored the necessity for targeted economic development strategies, especially in lower-wage regions. STEM jobs maintained a wage premium over non-STEM roles across various area types, emphasizing the importance of promoting STEM education and training.

Tailored interventions, including targeted STEM education initiatives and economic development strategies, were recommended to address regional disparities and stimulate job creation. Promoting higher education in STEM fields and aligning policies with analytical findings were advocated to support balanced economic growth and workforce development.

While the assumption of higher wages in STEM jobs was upheld, the analysis challenged the notion that employment size strongly influences wage levels, instead identifying industry-specific economic conditions and skill demands as primary determinants. This evolution of assumptions highlights the need for tailored interventions to address regional disparities and foster inclusive growth.




While the assumption that STEM jobs offer higher wages than non-STEM jobs was consistently upheld, the analysis challenged the assumption that employment size strongly influences wage levels. Instead, other factors such as industry-specific economic conditions and skill demands were found to be primary determinants of wage levels. As a result, the evolution of assumptions emphasizes the need for targeted interventions tailored to address regional disparities and promote inclusive growth across all types of geographical regions.

## THE DATA

The dataset utilized in this analysis comprises detailed wage information relevant to STEM occupations, enriched with attributes including Area Information, Industry Details, and Employment Statistics. This dataset was sourced to deepen the understanding of employment trends in STEM fields, focusing on elements like educational requirements and wage types. To ensure data integrity, extensive cleaning was necessary, particularly due to significant null values in several fields, which were methodically removed.

The relevance of this data lies in its ability to provide an extensive overview of wage distribution across educational levels within STEM, which is essential for nuanced analysis. The data preparation involved using Alteryx tools to refine and format the dataset, addressing challenges like missing data and deciding on the granularity of occupational data. The rationale behind these methodological choices, such as focusing on detailed SOC levels and excluding incomplete records, was driven by the need for precision in categorizing and analyzing employment trends accurately, crucial for effective policy-making.

## THE ANALYSIS

Assumptions regarding the dataset's completeness and the exclusion of incomplete wage data entries were crucial. These assumptions were later validated as analyses yielded consistent and reliable insights, confirming the effectiveness of the data preparation strategies.

Data blending was key in distinguishing between STEM and Non-STEM occupations, using a scalable method for large dataset categorization. This setup facilitated a comprehensive analysis across various geographical and industrial segments. Significant efforts were made to maintain data quality by filtering out rows with null values in key columns, which helped in conducting an unbiased wage differential analysis. Predictive models like Linear Regression, Decision Trees, and Random Forest were employed to analyze relationships between wages and economic indicators, enhancing the understanding of wage trends in STEM across diverse Metropolitan Statistical Areas (MSAs).

This structured approach provided clear, actionable insights, aiding in strategic decision-making and demonstrating the importance of meticulous data handling and analysis in understanding workforce dynamics in STEM fields.

The analytical journey began with a strategic focus on STEM occupations, driven by initial findings of significant variability in wage data and employment trends across regions and sectors. Utilizing Alteryx for data blending and advanced analytics platforms for statistical modeling, the team tackled extensive datasets, initiating the process with data cleansing and preparation. This included removing fields with high null values and creating lookup datasets for accurate STEM versus NON-STEM categorization.

A key milestone was the integration of cross-industry data, which expanded the team's understanding of STEM roles across different sectors, and the application of detailed SOC levels that enhanced the granularity of the analysis. Breakthroughs in predictive modeling were achieved by leveraging economic indicators such as inflation and GDP growth rates, providing new insights into wage dynamics under various economic conditions.

In the descriptive phase, the team dissected detailed characteristics and trends within STEM jobs, exploring wage distributions, employment statistics, and educational requirements across Metropolitan Statistical Areas (MSAs). Detailed SOC occupation levels were employed to gain a nuanced understanding of specific skill demands and wage discrepancies within STEM fields, laying the groundwork for diagnosing underlying patterns.

The predictive phase utilized machine learning models like Linear Regression, Decision Trees, and Random Forests to forecast wage differentials between STEM and NON-STEM occupations. This was pivotal in understanding how factors such as educational levels, economic indicators, and employment figures interacted to influence wage trends. The models incorporated adjustments for economic stability and utilized historical data to project future wage trends, providing actionable insights for strategic planning.

Building on the predictive insights, the prescriptive phase suggested specific policy interventions and educational initiatives to address disparities in wage trends and employment opportunities within STEM fields. Recommendations included targeted workforce development programs and adjustments in educational curricula to align with the evolving demands of the STEM labor market, aiming to enhance both employment and wage outcomes in these critical sectors.

(Please Refer to the Figures 1 to 21 from the appendix)

## FINDINGS AND RECOMMENDATIONS

The data analysis highlights the significant economic value attributed to STEM (Science, Technology, Engineering, Mathematics) jobs, consistently demonstrating higher wages across all geographical areas. This underscores the critical need for promoting STEM education and training to meet the demand for high-skilled labor. However, there are notable regional disparities in both employment levels and wage levels across metropolitan areas, indicating the necessity for targeted economic development strategies to stimulate job creation and wage growth, especially in lower-wage regions.

STEM jobs maintain a consistent premium in wages compared to non-STEM jobs, emphasizing the economic importance of STEM skills. This trend holds true across various area types, suggesting a nationwide demand for STEM expertise. The analysis also underscores the positive correlation between educational attainment and wage levels, highlighting the significance of higher education in securing high-paying STEM occupations.

Regional disparities in employment and wage levels necessitate tailored interventions to address specific needs. Targeted STEM education initiatives should be implemented, particularly in regions with disproportionately low STEM employment, to ensure a skilled workforce capable of accessing high-paying job opportunities. Furthermore, developing targeted economic development strategies in lower-wage regions is crucial to stimulating job creation and wage growth. These strategies should focus on enhancing access to education, vocational training, and industry diversification to foster sustainable economic growth.

Promoting higher education in STEM fields through scholarships, grants, and incentives can further equip individuals with the skills needed for lucrative STEM careers. Additionally, aligning policies with the findings of the analysis is essential to support both STEM and non-STEM sectors, ensuring a balanced approach to economic growth and workforce development.

While the assumption that STEM jobs offer higher wages than non-STEM jobs was consistently upheld, the analysis challenged the assumption that employment size strongly influences wage levels. Instead, other factors such as industry-specific economic conditions and skill demands were found to be primary determinants of wage levels. As a result, the evolution of assumptions emphasizes the need for targeted interventions tailored to address regional disparities and promote inclusive growth across all types of geographical regions.

## CONCLUSION

The analysis of wage dynamics within STEM and NON-STEM holistically and STEM occupations across Metropolitan Statistical Areas (MSAs) has uncovered several important insights. Firstly, STEM roles consistently command higher wages than NON-STEM positions across all regions, highlighting the substantial economic value of STEM skills. Furthermore, there is a positive correlation between higher educational attainment and increased wage levels in STEM fields, emphasizing the importance of advanced education in securing higher-paying jobs. Notably, there are significant regional disparities in employment levels and wage rates, indicating the need for targeted economic development strategies, especially in lower-wage areas.

The project evolved from an initial broad dataset analysis to a detailed examination of variables affecting STEM wages, employing advanced analytical tools and methodologies. We overcame significant data challenges, refined our methods, and applied complex models to not only understand but also forecast future trends, providing strategic insights into the STEM labor market.

Investing in STEM education and training emerges as crucial for securing higher-paying jobs and enhancing economic prosperity. Recommendations focus on targeting educational programs in regions with low STEM employment and adjusting policies to support comprehensive regional development. The initial assumptions about the uniform impact of economic factors and the correlation between employment size and wages were refined through our research, revealing that wages are influenced more significantly by industry-specific conditions and specialized skills than by workforce size.

## REFERENCES


*Current US inflation rates: 2000-2024* (2024) *US Inflation Calculator | Easily calculate how the buying power of the U.S. dollar has changed from 1913 to 2023. Get inflation rates and U.S. inflation news.* Available at: https://www.usinflationcalculator.com/inflation/current-inflation-rates/ (Accessed: 06 May 2024).

Deming, D. and Noray, K. (2018) *Stem careers and the changing skill requirements of work* [Preprint]. doi:10.3386/w25065.

Harmon, C., Oosterbeek, H. and Walker, I. (2003) 'The returns to education: Microeconomics', *Journal of Economic Surveys*, 17(2), pp. 115–156. doi:10.1111/1467-6419.00191.

Ghazi, S. (2023) *Understanding the gaps in the US stem labor market, Oxford Economics*. Available at: https://www.oxfordeconomics.com/resource/understanding-the-gaps-in-the-us-stem-labor-market/ (Accessed: 02 May 2024).

*{ indicator.label }* (no date) *IMF*. Available at: https://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD/USA (Accessed: 06 May 2024).

(No date) *The stem labor force of today: Scientists, engineers, and skilled technical workers | NSF - national science foundation*. Available at: https://ncses.nsf.gov/pubs/nsb20212/stem-labor-market-conditions-and-the-economy (Accessed: 01 May 2024).

(No date) *The stem labor force of today: Scientists, engineers, and skilled technical workers | NSF - national science foundation*. Available at: https://ncses.nsf.gov/pubs/nsb20212/stem-labor-market-conditions-and-the-economy (Accessed: 02 May 2024).

*Oes Home* (no date) *U.S. Bureau of Labor Statistics*. Available at: https://www.bls.gov/oes/ (Accessed: 01 May 2024).

# APPENDIX

# OVERALL ASSUMPTIONS AND METHODOLOGICAL CHOICES

# Assumptions and Methodological Choices

## Assumptions

**Utilizing Detailed SOC Occupation Levels:**
- Assumes detailed SOC occupation levels provide accurate insights into STEM occupations for targeted interventions and policy responses.
- Presumes that nuanced dynamics of STEM occupations can be precisely captured at this detailed level, essential for informed decision-making.

**Adjusting Annual Mean Wages for Economic Indicators:**
- Assumes that adjusting annual mean wages for inflation and GDP growth rate enhances predictive validity by capturing broader macroeconomic influences on wage dynamics.
- Presumes that changes in economic indicators reflect underlying economic factors that impact wages, such as labor market conditions and productivity growth.

**Focusing on MSAs for STEM Wage Analytics:**
- Assumes that focusing on MSAs provides robust datasets and detailed insights applicable to economic strategies.
- Presumes that MSAs represent core hubs of innovation and economic activity, making them strategically important for workforce development and policy planning.

**Focusing on High STEM Employment Areas in MSAs:**
- Assumes that targeting high STEM employment areas in MSAs strategically reveals key trends for workforce development.
- Presumes that these areas serve as models for emerging sectors and draw substantial investments, influencing regional job markets.

## Methodological Choices

**Removing Rows with Null Values in Wage Columns:**
- Methodological choice to ensure only complete and reliable data are used for wage differential analysis.
- Avoids bias in the analysis by excluding incomplete or unreliable data points.

**Filtering "I_GROUP" Field Using 4-Digit NAICS Classification:**
- Methodological choice to achieve balanced granularity for meaningful comparisons across industries.
- Ensures alignment with industry standards and facilitates manageable complexity in the analysis.

**Excluding "Total" Entries from the Dataset:**
- Methodological choice to enhance precision by preventing double counting of employment and wage data.
- Ensures each data point specifically represents an individual occupation, providing granularity for accurate analysis.

**Including Cross-Industry Data:**
- Methodological choice to ensure a comprehensive view of employment trends, particularly in STEM occupations.
- Acknowledges that many STEM occupations transcend traditional industry boundaries and intersect multiple sectors.

**Choosing "Detailed" SOC Levels:**
- Methodological choice to ensure accurate identification and categorization of occupations, minimizing misclassifications.
- Presumes that detailed SOC levels provide the granularity needed for precise analysis of specific STEM and non-STEM roles.

**Utilizing Both Mean and Median Metrics:**
- Methodological choice to provide a balanced wage analysis approach, aiding decision-making.
- Recognizes that both metrics offer complementary insights into wage dynamics, capturing different aspects of the wage distribution.

# DIARY ENTRIES

## 1 My MN5816 Development Diary

| Date | Code | My Diary Entries |
|---|---|---|
| 15-03-2024 | D | • Selected the topic-The Effect of Interest Rate and Government stimulus on USA Inflation rates over the past 10 years. |
| 10-04-2024 | A | • Analysed the data set for inflation, Government stimuli, interest rates and the depth of scholarly understanding in economics specifically. |
| 30-04-2024 | D | • Considering the small size of the dataset which could limit the data analysis and most importantly the topic being very extensive requiring in-depth knowledge in economics decided to change the data set. |
| 01-05-2024 | C | • Selected USA wage labour dataset as the dataset. Primary dataset being the United states OWES data for all sectors and all areas. Additional data sets of STEM and Educational Qualification can be used. |
| 02-05-2024 | A | • Analysed the dataset to confirm its feasibility and alignment with the project requirement and moved ahead with the dataset. |
| 02-05-2024 | D | • Decided to move ahead with data set to analyse the Wage for STEM occupations based on area,industry,payment, etc and make required predictions. |
| 03-05-2024 | A & D | • Its important to plan the whole process since the dataset is big in size. After carefully analysing the dataset, it is suggested to work on a hoslistic descriptive analytics first to give an overview of the STEM occupation analytics, then to move on to granular descriptive analytics filtering based on, for instance, area or state. This should make the predictive modeling as precise as it can be. Another point to be noted is that before sampling for the prediction, descriptive analytics has to be completed because the volume of data supports descriptive analytics in terms of reliability. |
| 03-05-2024 | A | • The plan for the predictive model is to forecast the wages( annually or hourly, whichever is reasonable and relevant). The initial variables can be supply/demand in terms of employed numbers, and also some macro economic factors like Inflation. However, both holistic and granular descriptive analytics must form the base for the predictions to be made (predictors, target variable and so on) |
| 03-05-2024 | A | • Assumed that the 2023 BLS Occupational Employment and Wage Statistics (OEWS) data would provide detailed insights into wage dynamics across STEM occupations. |

| Date | Code | My Diary Entries |
|---|---|---|
| 03-05-2024 | D | • Decided to focus on STEM occupations in Metropolitan Statistical Areas (MSAs) as they offer a significant concentration of specialized skills and industries. |
| 03-05-2024 | M | • Identified key attributes required for analysis: NAICS, SOC codes, wage and employment data, location quotients. |
| 03-05-2024 | C | • Encountered significant missing values in key wage columns like `Hourly Mean` and `Annual Mean`. |
| 03-05-2024 | D | • Decided to exclude rows with null values in wage columns to maintain data integrity. |
| 03-05-2024 | A | • Assumed that removing incomplete records would not significantly skew the analysis. |
| 03-05-2024 | A | • Assumed that detailed SOC occupation levels would yield the most accurate insights into STEM wage dynamics. |
| 03-05-2024 | D | • Created a lookup dataset to categorize occupations as STEM or NON-STEM based on SOC codes. |
| 03-05-2024 | M | • Built a scalable categorization system to easily filter and analyze STEM occupations. |
| 03-05-2024 | C | • Difficulty distinguishing STEM from NON-STEM roles in cross-industry data |
| 03-05-2024 | D | • Decided to include cross-industry data and use detailed SOC levels for accurate categorization. |
| 03-05-2024 | A | • Assumed that this approach would minimize misclassification errors. |
| 03-05-2024 | A | • Assumed that focusing on STEM occupations within MSAs would provide a representative sample of the STEM workforce. |
| 03-05-2024 | D | • Filtered dataset to include only data points within MSAs and categorized by area types. |
| 03-05-2024 | C | • Data cleansing required due to inconsistencies in area titles. |
| 03-05-2024 | D | • Created initial visualizations showing wage disparities between STEM and NON-STEM across MSAs. |

| Date | Code | My Diary Entries |
|---|---|---|
| **03-05-2024** | **M** | • Used heat maps and box plots to present clearer insights into regional wage trends. |
| **03-05-2024** | **A** | • Assumed that GDP growth and inflation would have differential impacts on STEM wages across regions. |
| **04-05-2024** | **D** | • Included economic indicators like GDP growth rate and inflation rate to adjust wage predictions. |
| **04-05-2024** | **M** | • Introduced Adjusted Annual Mean to reflect economic adjustments based on external indicators. |
| **04-05-2024** | **D** | • Developed a formula to calculate the Adjusted Annual Mean using inflation and GDP growth rates:<br>Adjusted Annual Mean = Annual Mean * (1 + (GDP Growth Rate - Inflation Rate)) |
| **04-05-2024** | **M** | • Estimated missing data points using state-level economic indicators. |
| **04-05-2024** | **C** | • Challenges in accurately predicting STEM wages due to nonlinear relationships between predictors. |
| **04-05-2024** | **D** | • Decided to use machine learning models like Decision Trees and Random Forests to handle complex data relationships. |
| **04-05-2024** | **A** | • Assumed that employment numbers (TOT_EMP) would be a strong predictor of wage levels due to supply and demand dynamics. |
| **04-05-2024** | **D** | • Implemented Random Forest models to predict Adjusted Annual Mean based on economic and occupational predictors. |
| **04-05-2024** | **M** | • Achieved significant predictive accuracy improvements, with Random Forest models outperforming linear regression. |
| **04-05-2024** | **C** | • Difficulty standardizing educational requirement data across different occupational groups. |
| **04-05-2024** | **M** | • Used the Education field to categorize and analyze wage differentials by educational level. |
| **04-05-2024** | **D** | • Included Typical Entry-Level Educational Requirement as a key predictor in predictive models. |
| **04-05-2024** | **D** | • Developed detailed descriptive analytics focusing on STEM wage distributions across MSAs. |

| Date | Code | My Diary Entries |
|---|---|---|
| **05-05-2024** | **D** | • Finalized the predictive models for Adjusted Annual Mean with Random Forest and Decision Tree approaches. |
| **05-05-2024** | **M** | • Compiled strategic recommendations for policymakers and educational institutions based on findings. |
| | | • |

***Declaration of AI Tool Usage***

This declaration must be completed if you have used any AI tool in the completion of this assignment. This includes Quillbot, Grammarly, Paperpal, Elicit, Scispace, ChatGPT, Claude, Perplexity or any other tools of this kind.
There are many ethical and acceptable uses of AI tools that have been explained to you during lectures and workshops. If you have any doubt about your use of AI, then please contact your tutor to seek advice.
There is, however, one significant 'unethical' use that can result in an academic misconduct investigation. You must not use AI-generated text, whether this is a single phrase, sentence, or paragraph, presented as your own work in your submission. This also includes allowing any AI tool (e.g. Grammarly) to rephrase of your work, because it will no longer be your work.

| ***AI Tool*** | ***Subscription (YES/NO)*** | ***How and for What purpose was this tool used?*** |
|---|---|---|
| ChatGPT | No | Used for research and resolving Alteryx errors. Assisted in understanding data cleansing techniques and identifying best practices for predictive modelling. |
| Copilot | Yes (provided by RHUL) | Assisted in debugging and optimizing analysis code, particularly in refining predictive analytics workflows. |
| Scispace | No | Assisted with document management and version control, ensuring accurate tracking of report changes and facilitating seamless collaboration. |

## ALTERYX DATA VISUALIZATION

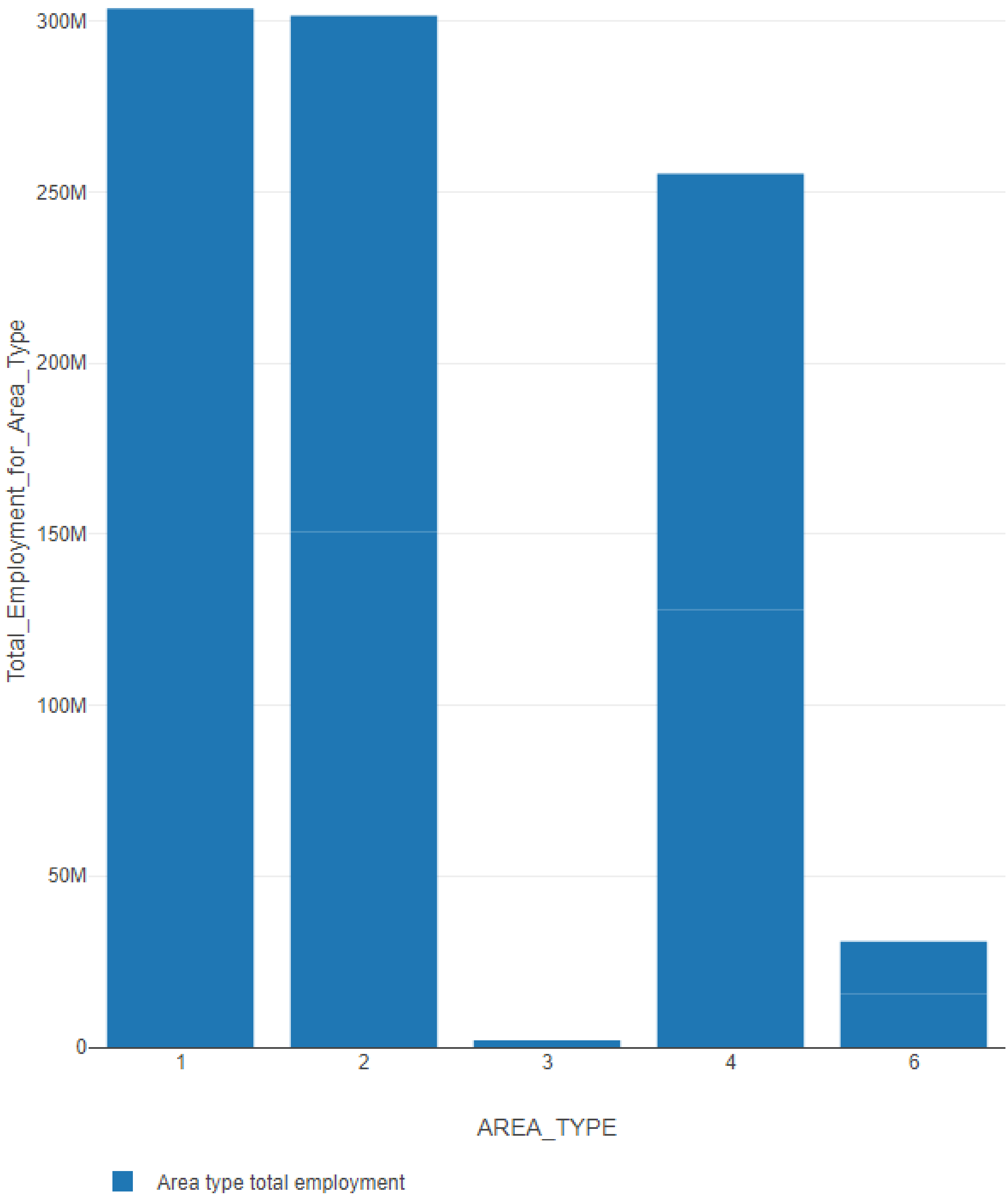



**Fig:1**

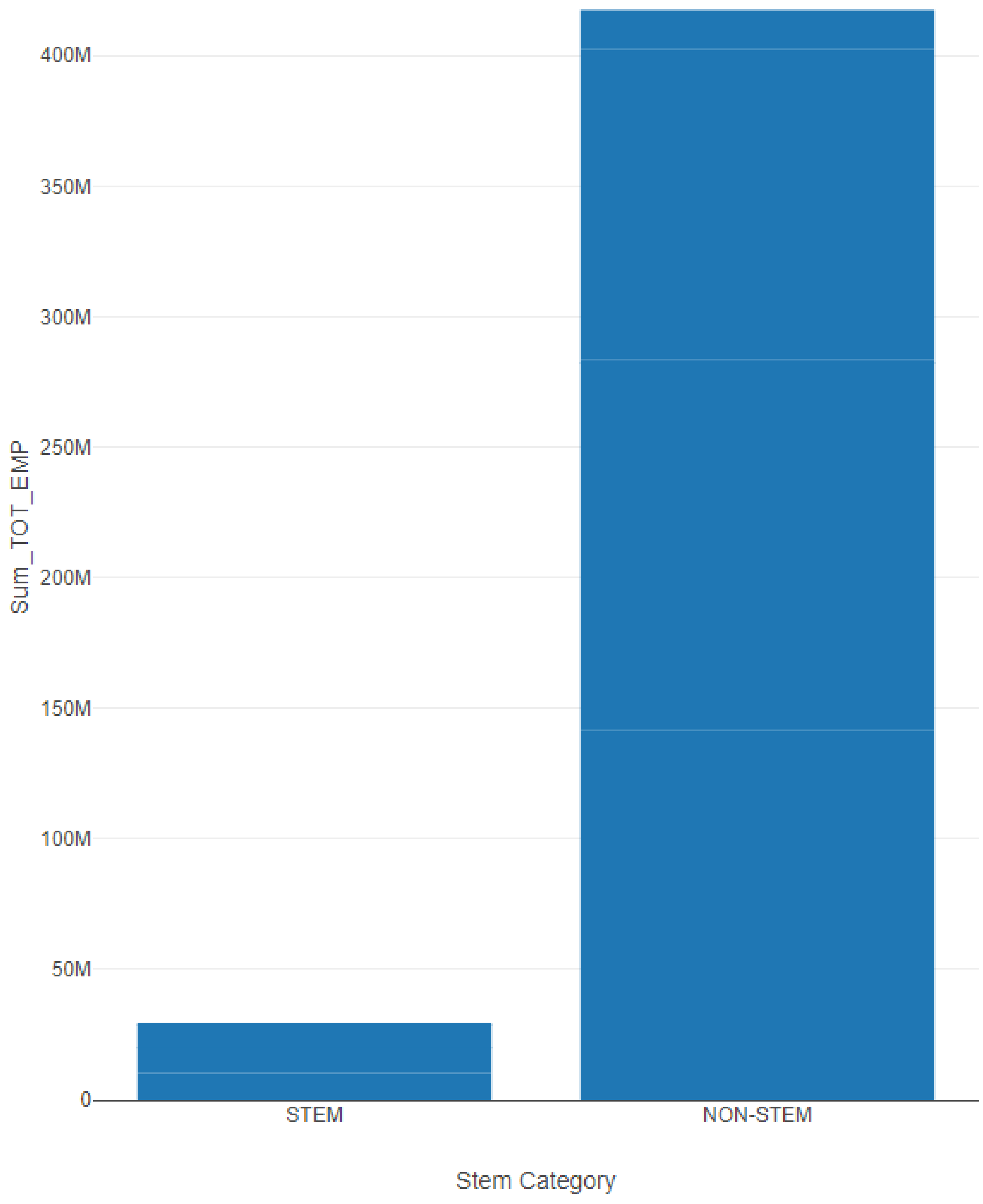

400M
350M
300M
250M
200M
150M
100M
50M
0
Sum_TOT_EMP
STEM
NON-STEM
Stem Category

**Fig:2**

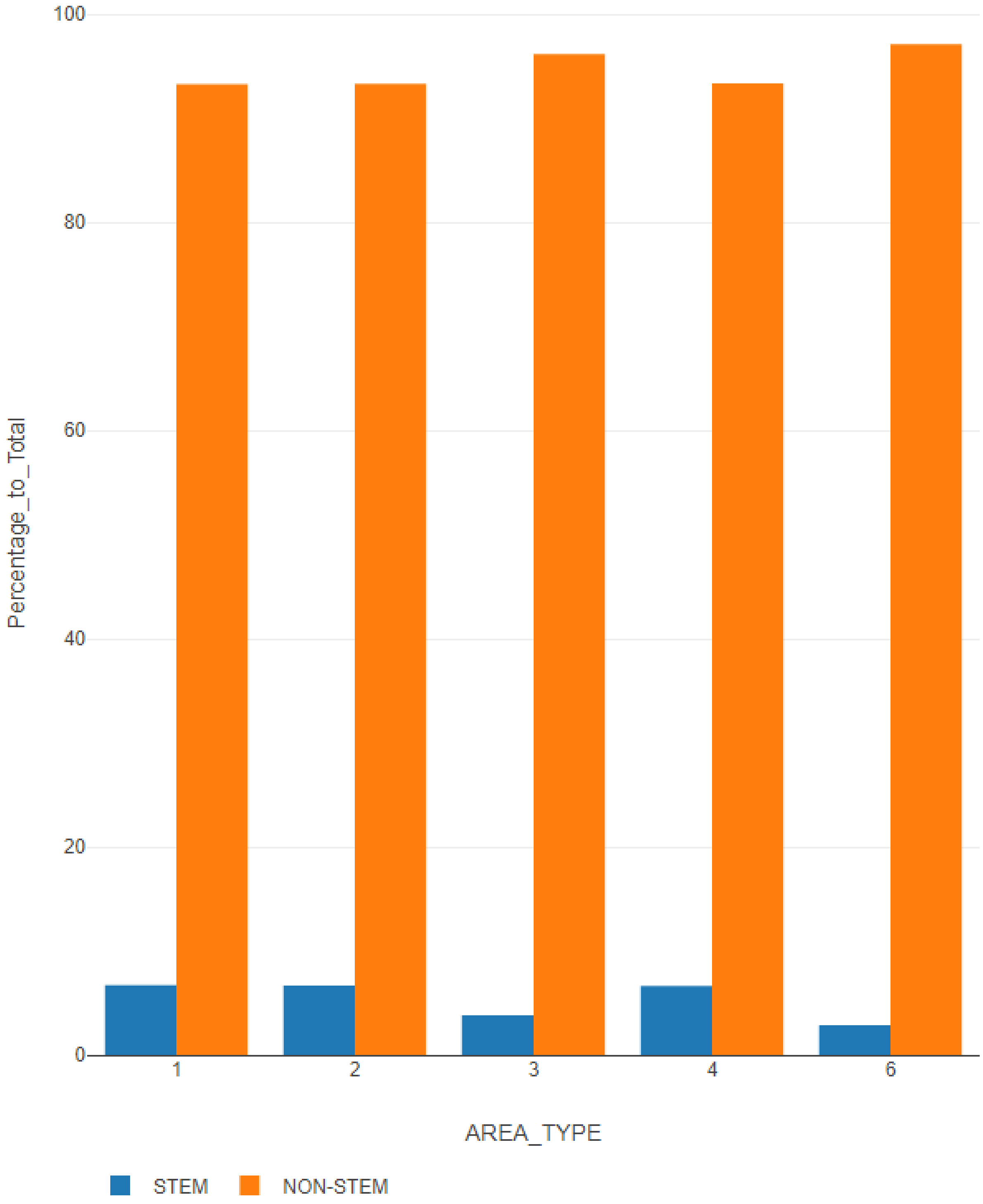

100
80
60
40
20
0
Percentage_to_Total
1
2
3
4
6
AREA_TYPE
STEM
NON-STEM

**Fig:3**

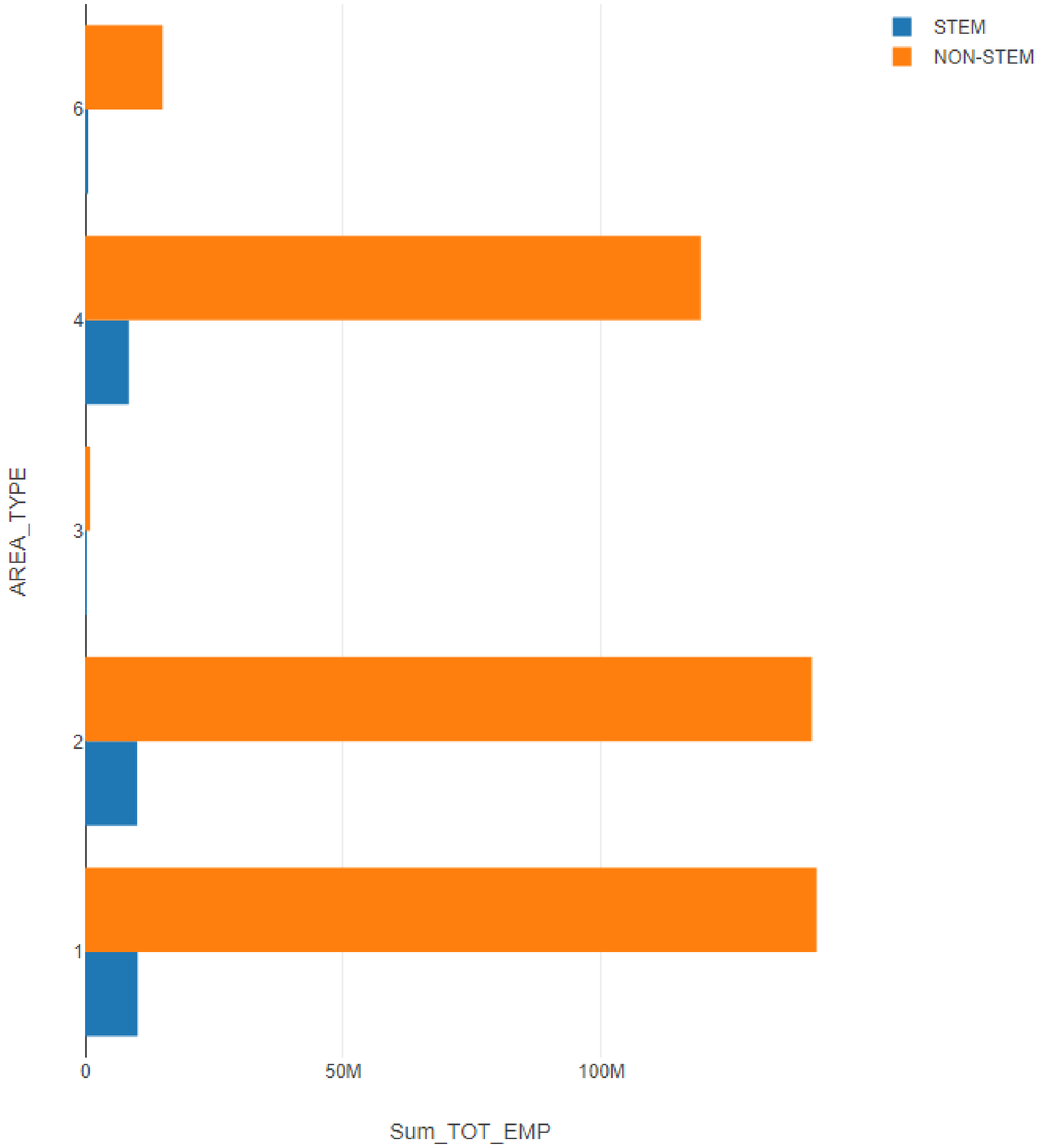



**Fig:4**

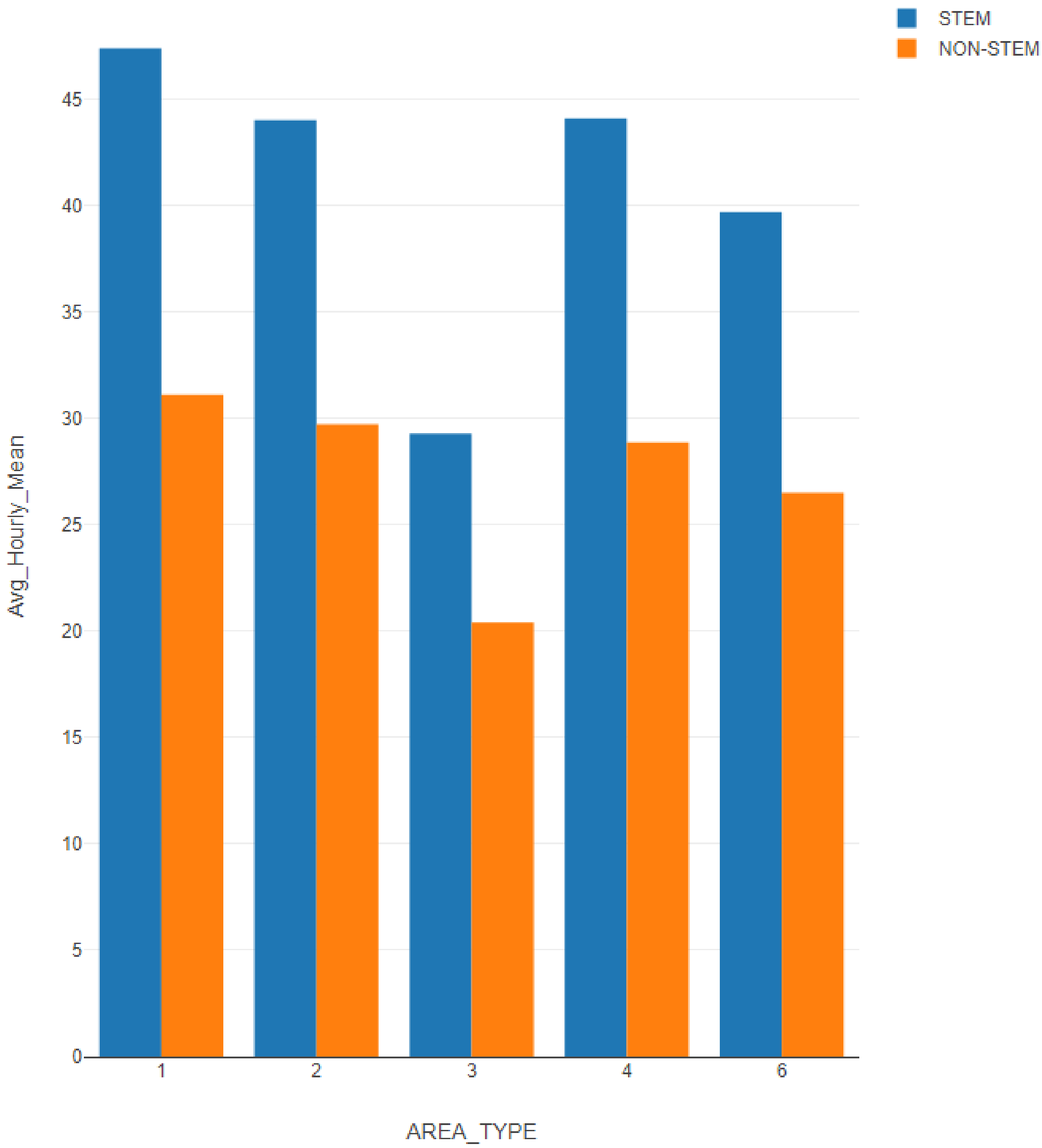

STEM
NON-STEM
45
40
35
30
25
20
15
10
5
0
Avg_Hourly_Mean
1
2
3
4
6
AREA_TYPE

**Fig:5**

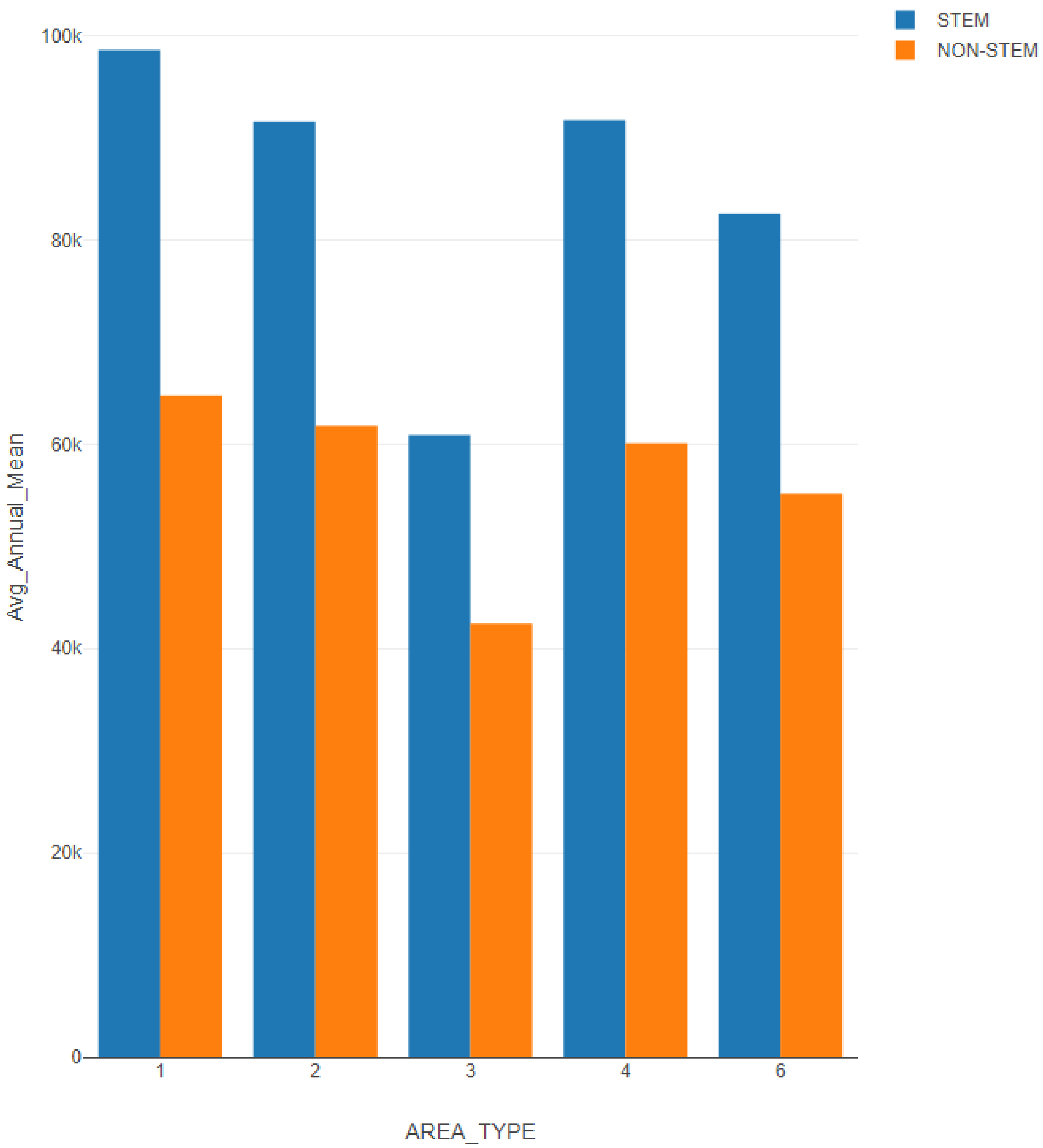
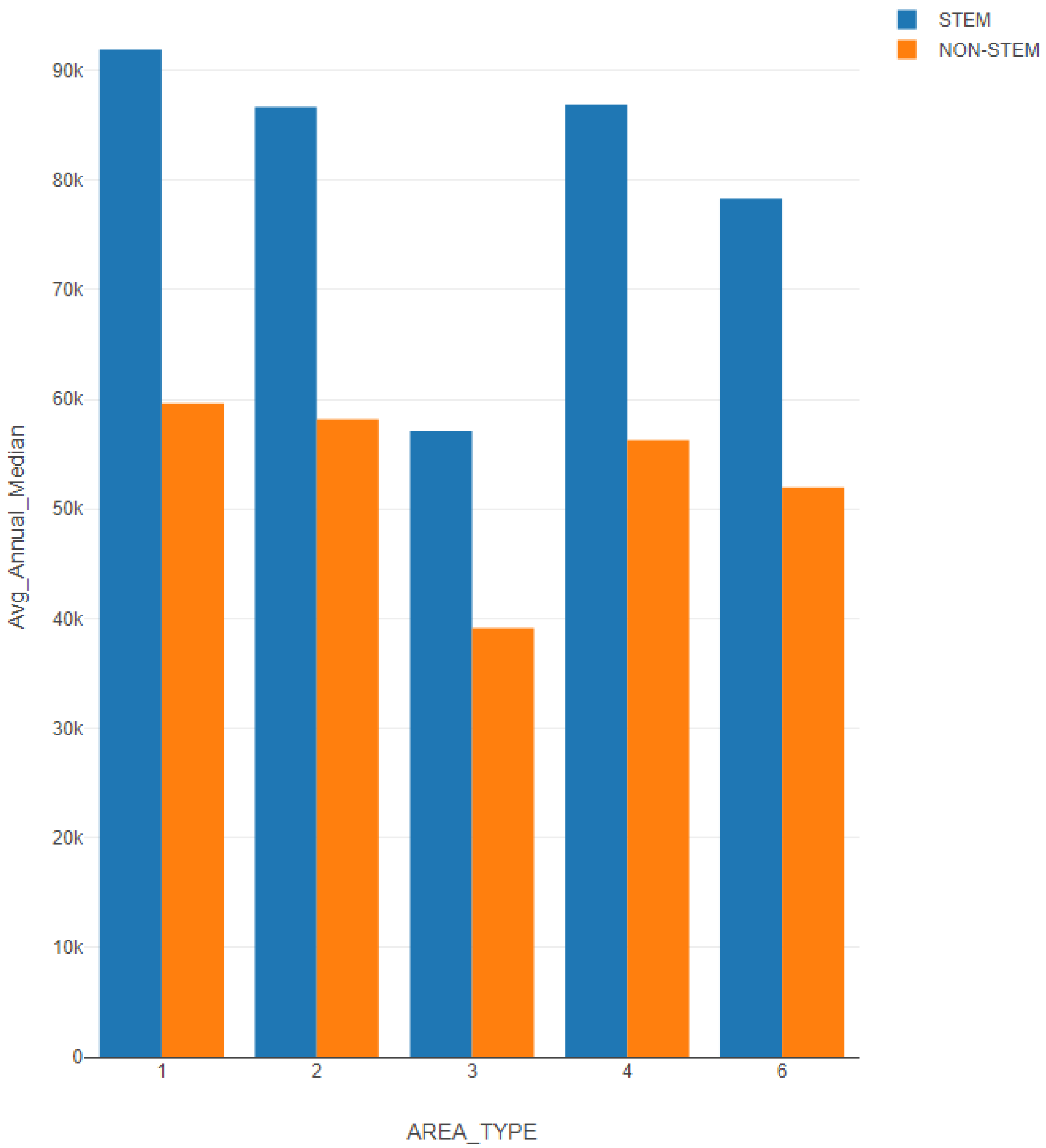
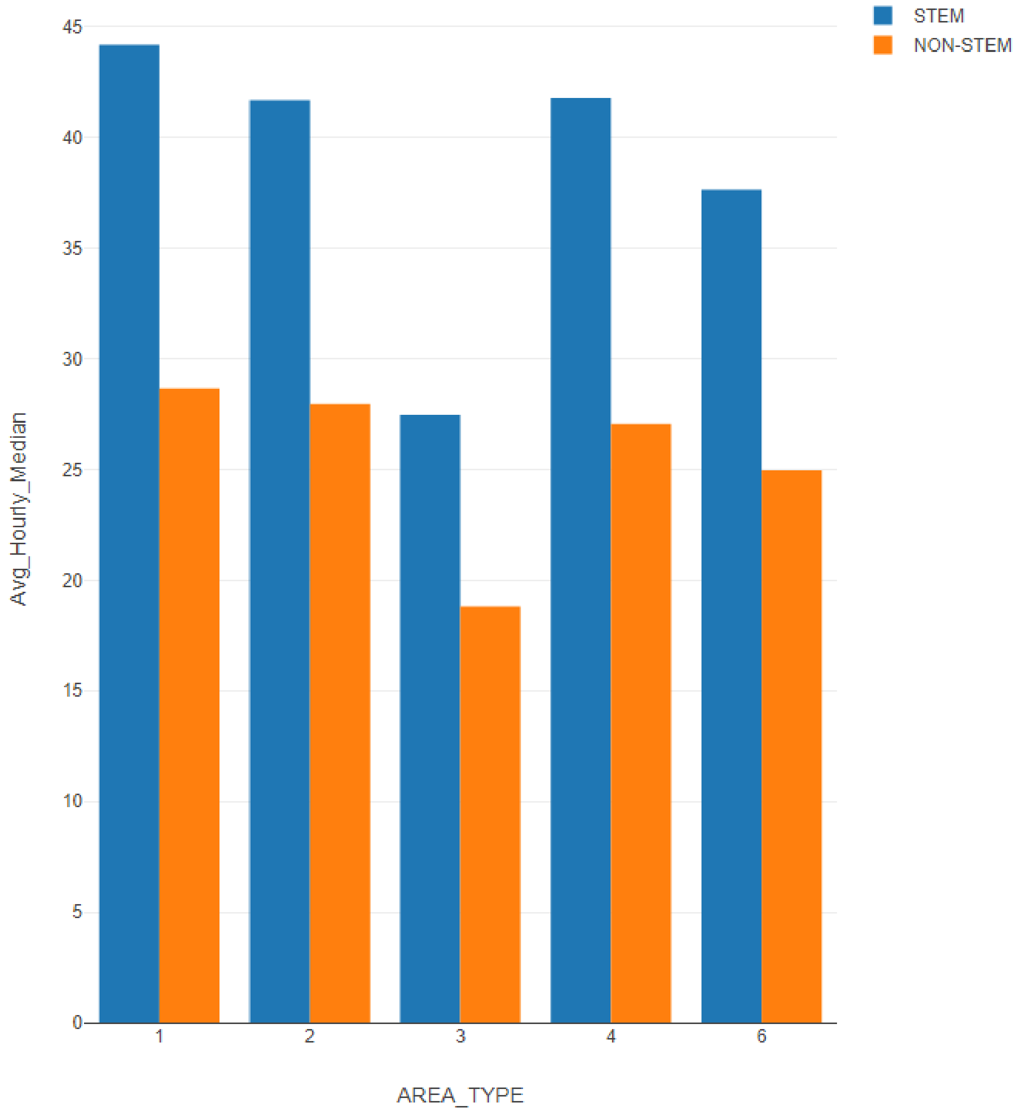

STEM
NON-STEM
100k
80k
60k
40k
20k
0
Avg_Annual_Mean
1
2
3
4
6
AREA_TYPE

**Fig:6**



STEM
NON-STEM
90k
80k
70k
60k
50k
40k
30k
20k
10k
0
Avg_Annual_Median
1
2
3
4
6
AREA_TYPE


**Fig:7**



STEM
NON-STEM
Avg_Hourly_Median
45
40
35
30
25
20
15
10
5
0
1
2
3
4
6
AREA_TYPE


**Fig:8**

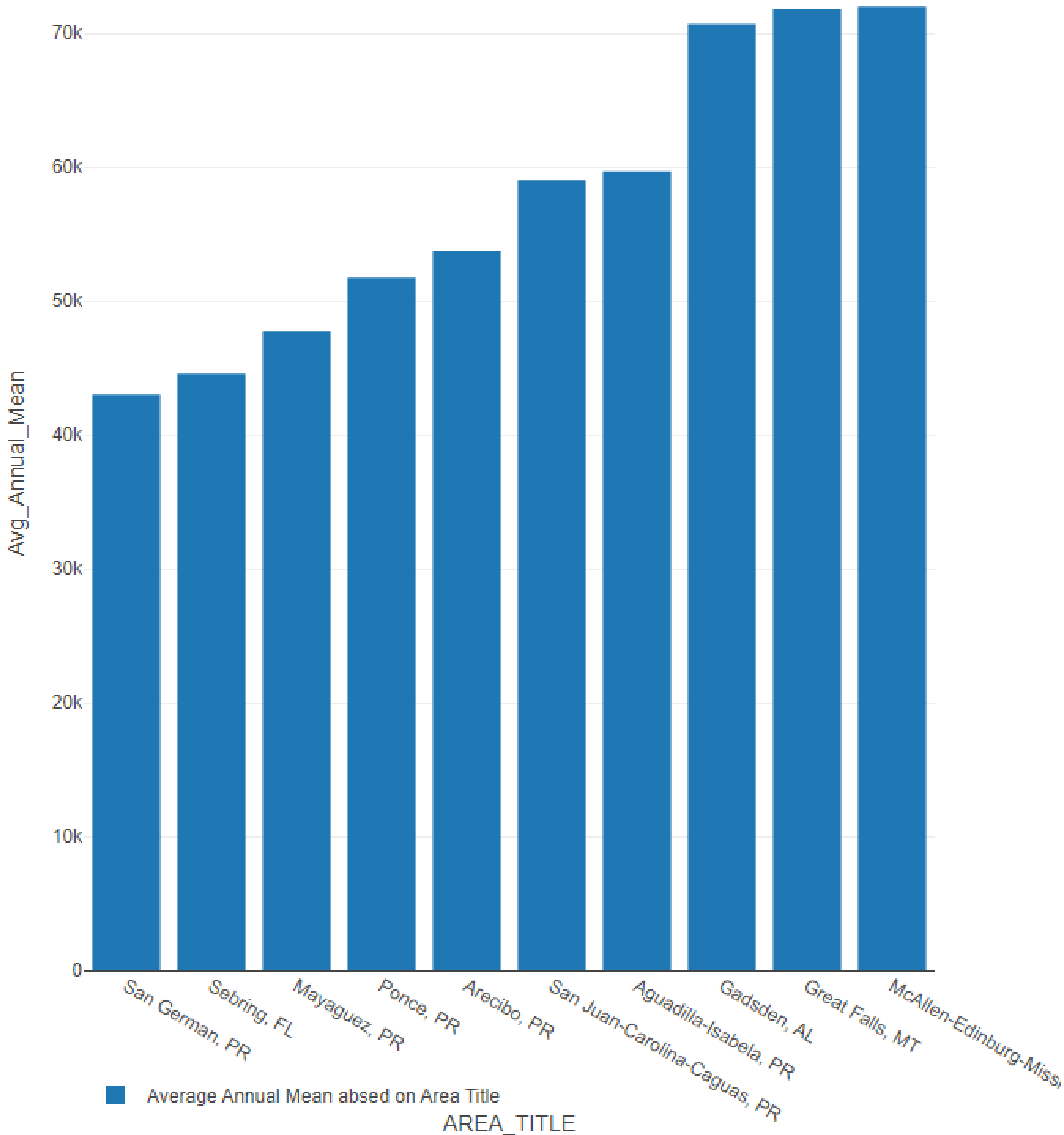

70k
60k
50k
40k
30k
20k
10k
0
Avg_Annual_Mean
San German, PR
Sebring, FL
Mayaguez, PR
Ponce, PR
Arecibo, PR
San Juan-Carolina-Caguas, PR
Aguadilla-Isabela, PR
Gadsden, AL
Great Falls, MT
McAllen-Edinburg-Miss,
Average Annual Mean absed on Area Title
AREA_TITLE

**Fig:9**

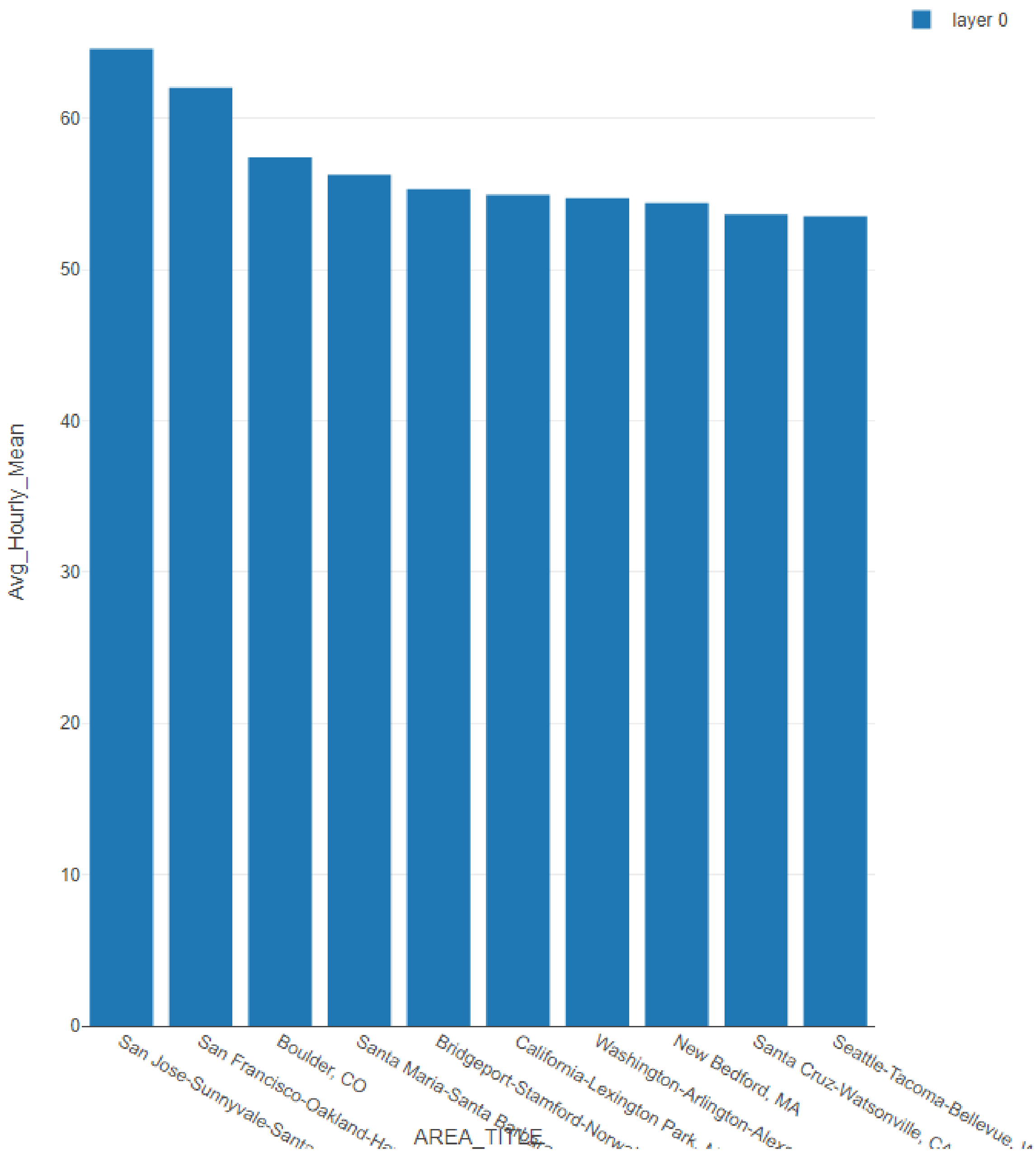

layer 0
Avg_Hourly_Mean
60
50
40
30
20
10
0
AREA_TITLE
Boulder, CO
New Bedford, MA

**Fig:10**

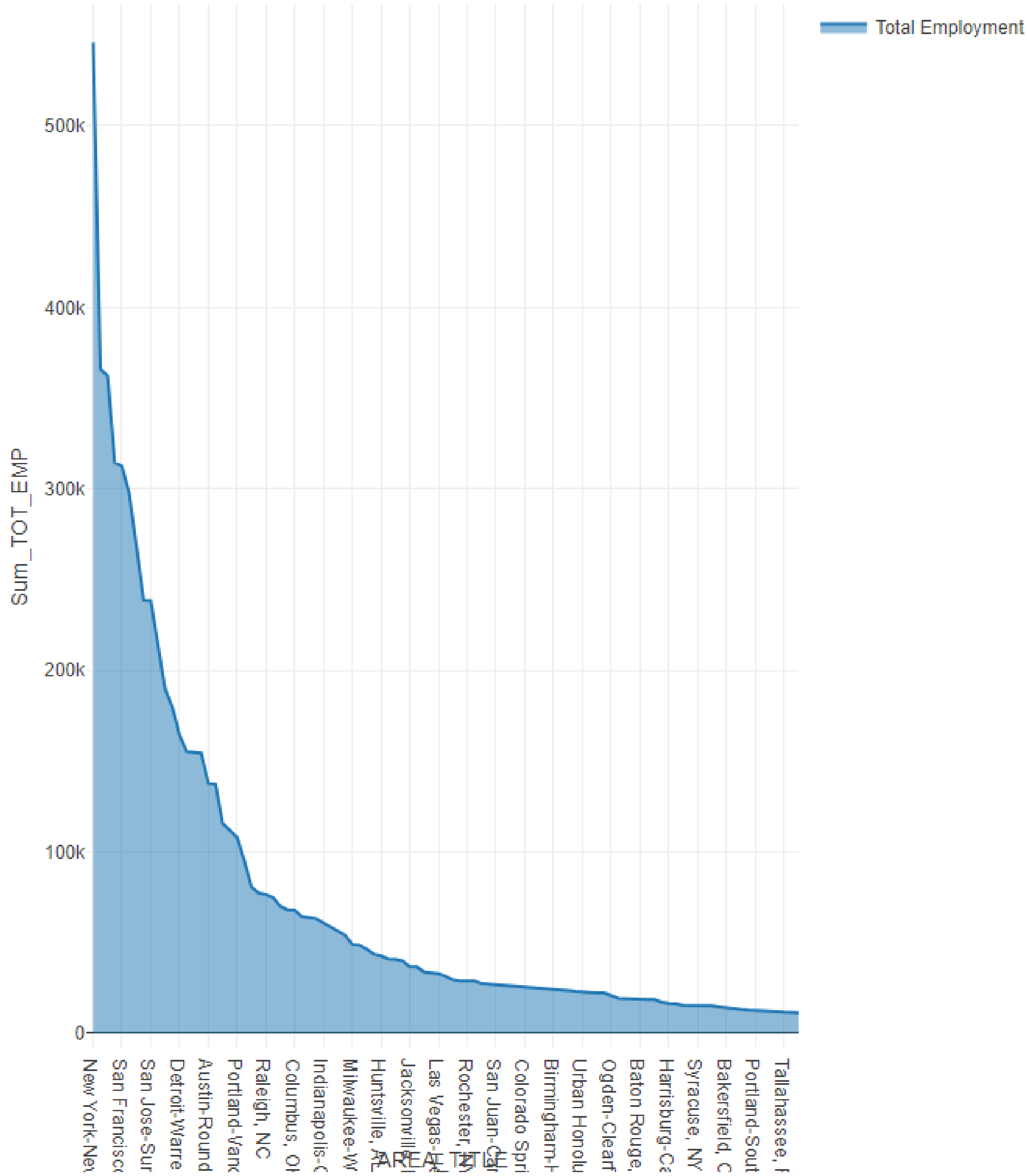

Total Employment
Sum_TOT_EMP
500k
400k
300k
200k
100k
0
AREA_TITLE
New York-New
San Francisco
San Jose-Sun
Detroit-Warre
Austin-Round
Portland-Vanc
Raleigh, NC
Columbus, OH
Indianapolis-C
Milwaukee-W
Huntsville, AL
Jacksonville
Las Vegas-
Rochester, N
San Juan-C
Colorado Spri
Birmingham-H
Urban Honolu
Ogden-Clearf
Baton Rouge,
Harrisburg-Ca
Syracuse, NY
Bakersfield, C
Portland-Sout
Tallahassee, F

**Fig:11**

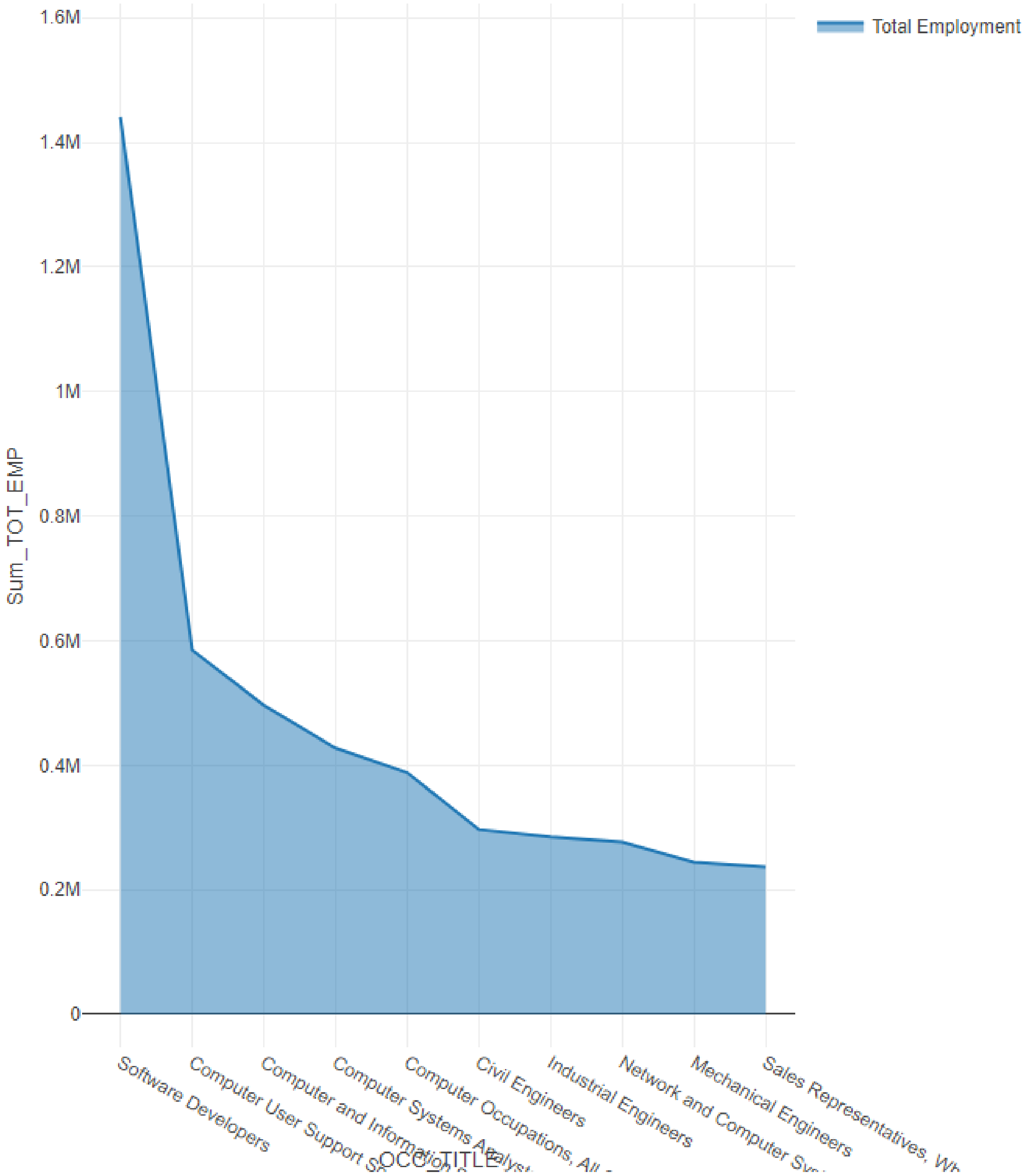
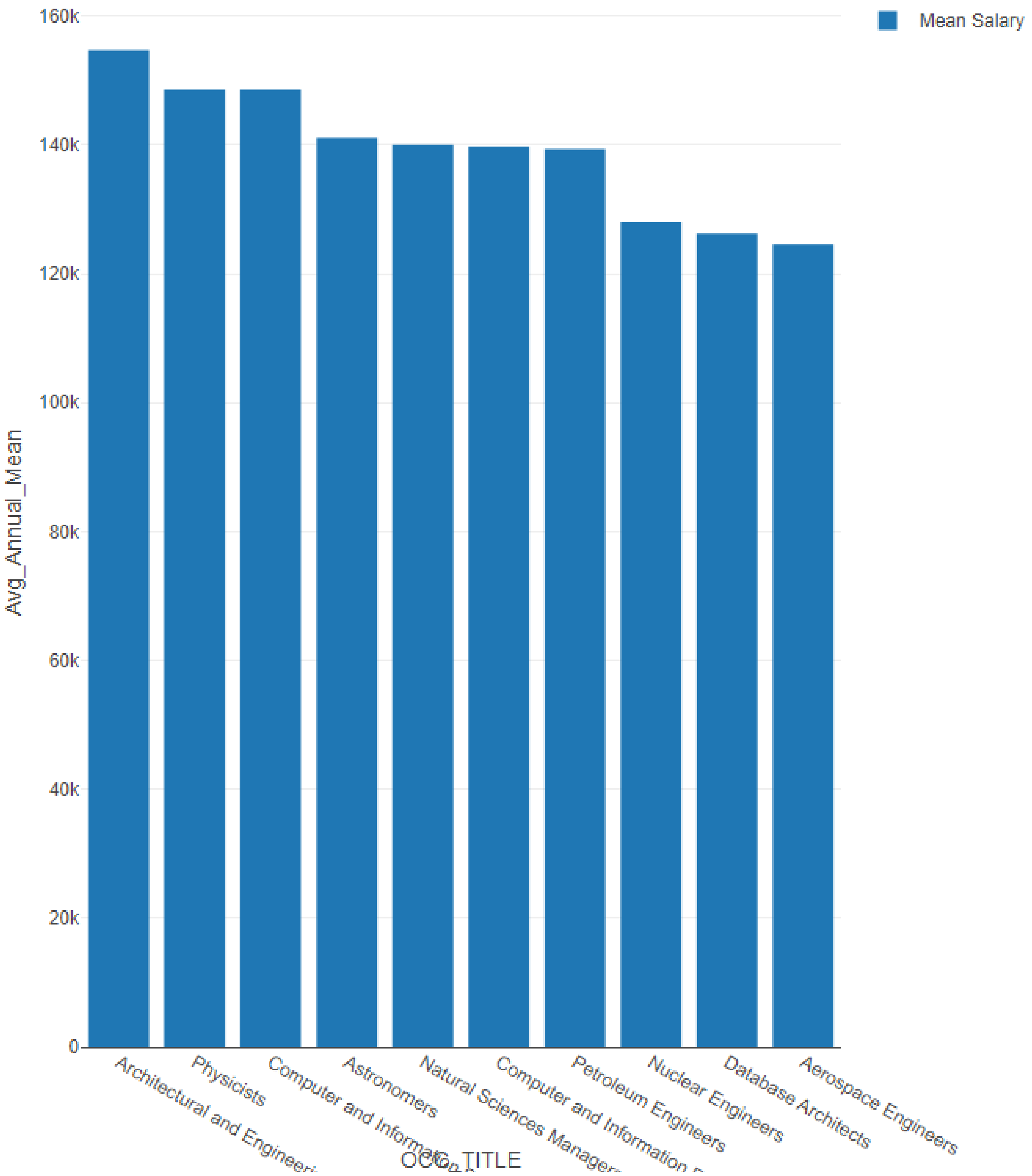
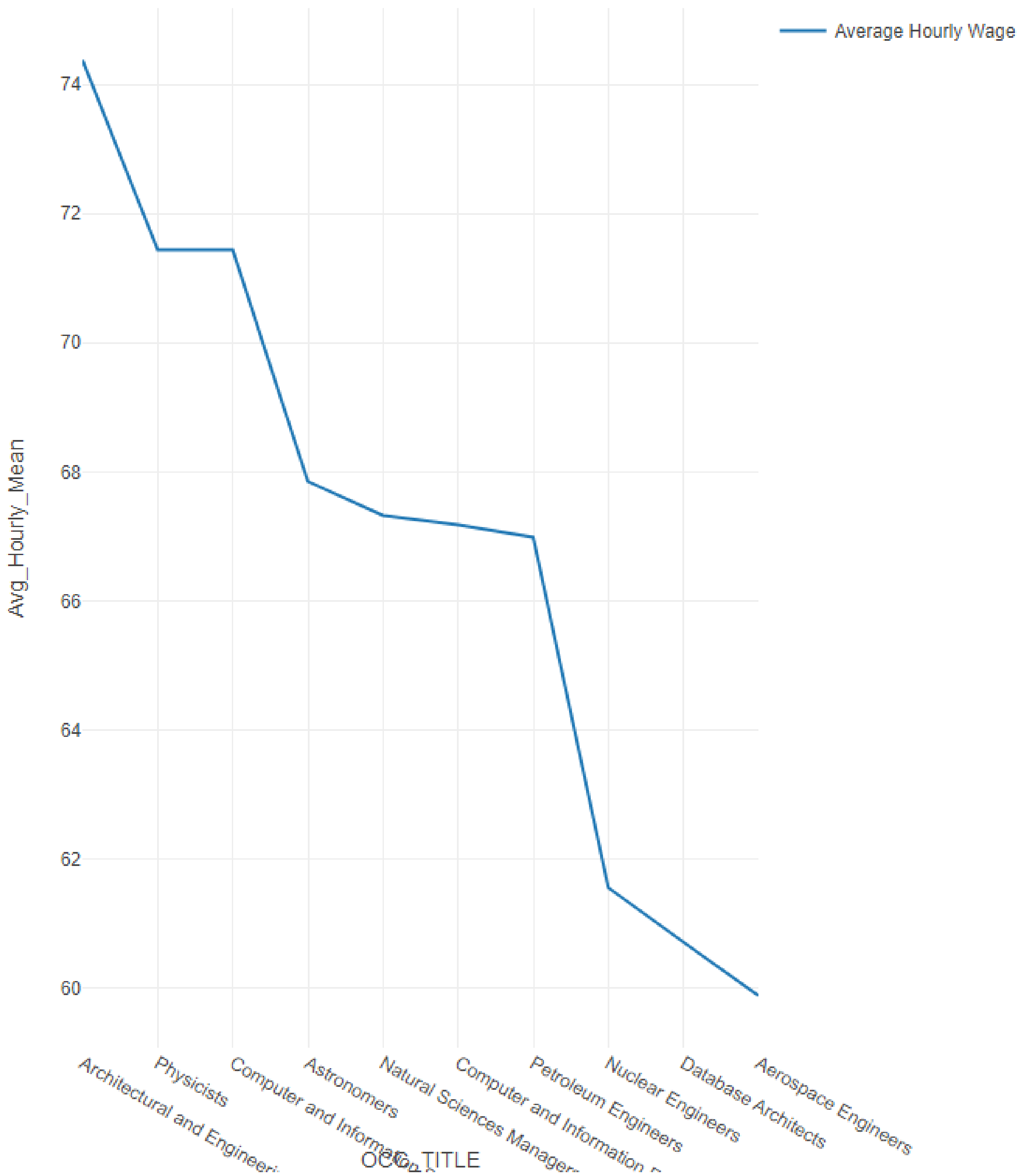

Total Employment
Sum_TOT_EMP
1.6M
1.4M
1.2M
1M
0.8M
0.6M
0.4M
0.2M
0
OCC_TITLE
Software Developers
Civil Engineers
Industrial Engineers
Mechanical Engineers

**Fig:12**



Mean Salary
160k
140k
120k
100k
80k
60k
40k
20k
0
Avg_Annual_Mean
Architectural and Engineeri...
Physicists
Computer and Informati...
Astronomers
Natural Sciences Manager...
Computer and Information R...
Petroleum Engineers
Nuclear Engineers
Database Architects
Aerospace Engineers
OCC_TITLE


**Fig:13**



Average Hourly Wage
Avg_Hourly_Mean
74
72
70
68
66
64
62
60
Architectural and Engineeri...
Physicists
Computer and Informatio...
Astronomers
Natural Sciences Manager...
Computer and Information R...
Petroleum Engineers
Nuclear Engineers
Database Architects
Aerospace Engineers
OCC_TITLE


**Fig:14**

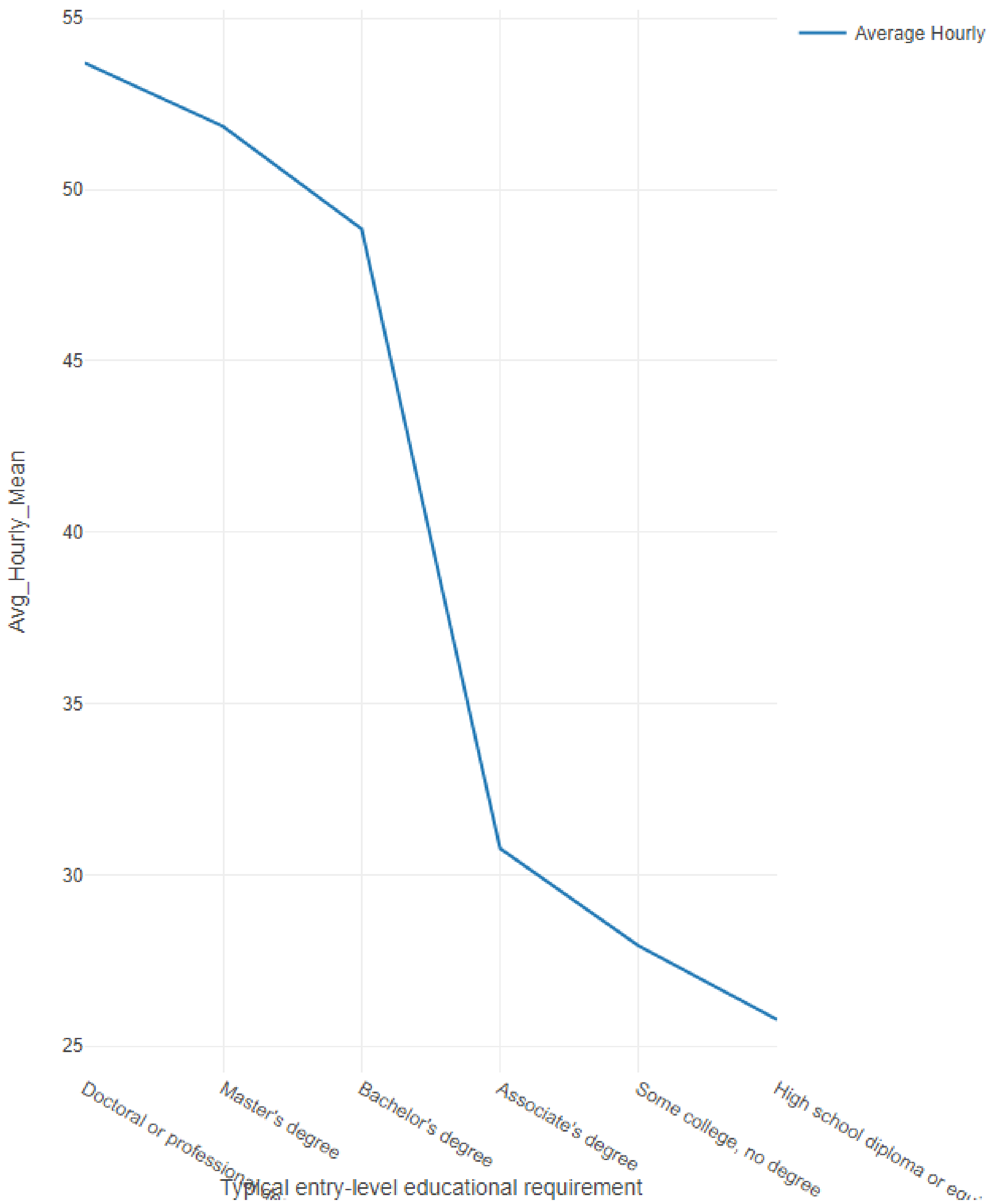

Average Hourly Mean
Avg_Hourly_Mean
55
50
45
40
35
30
25
Doctoral or professional...
Master's degree
Bachelor's degree
Associate's degree
Some college, no degree
High school diploma or eq...
Typical entry-level educational requirement

**Fig:15**

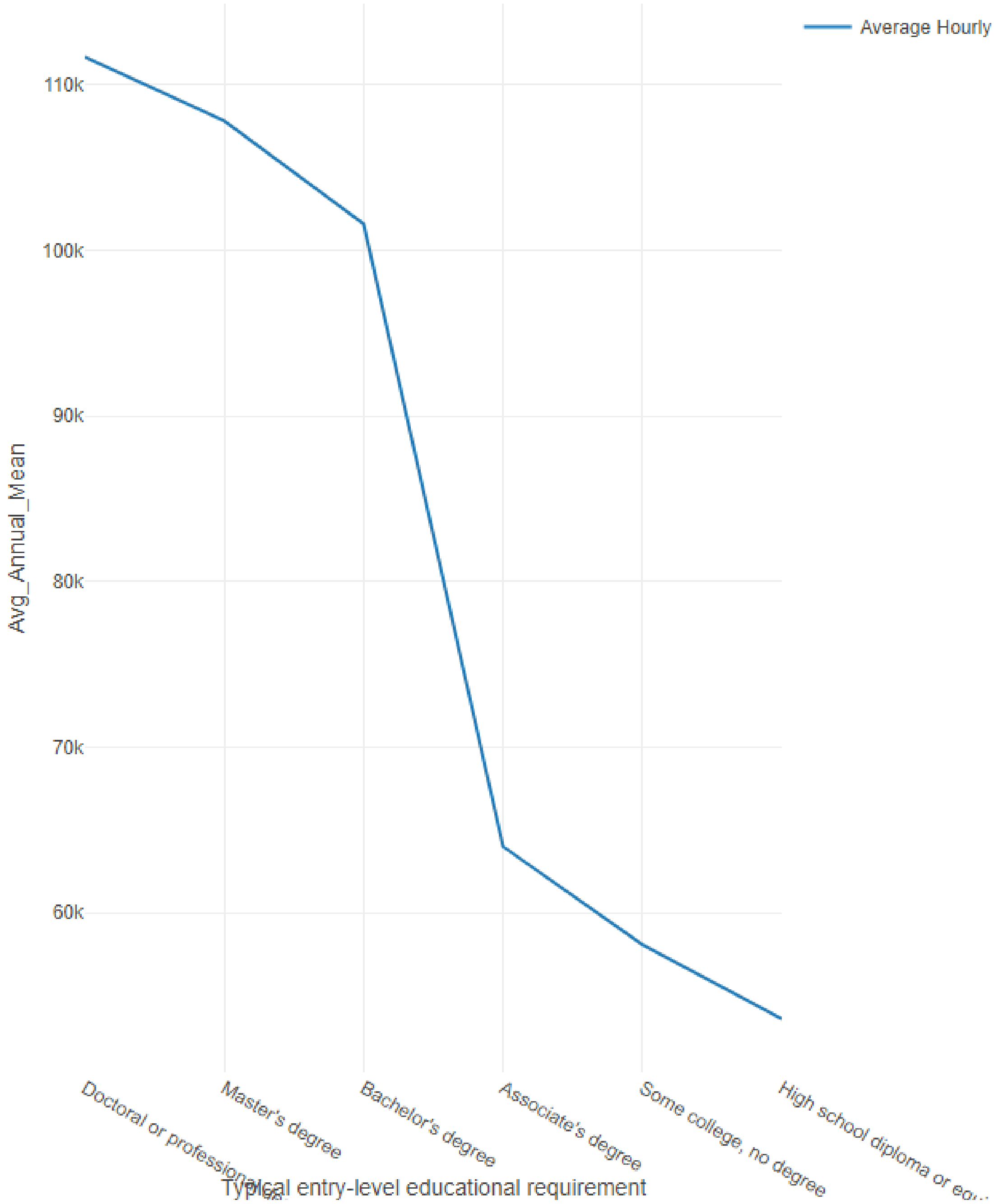

Average Hourly Mean
110k
100k
90k
80k
70k
60k
Avg_Annual_Mean
Doctoral or professional...
Master's degree
Bachelor's degree
Associate's degree
Some college, no degree
High school diploma or eq...
Typical entry-level educational requirement

**Fig:16**

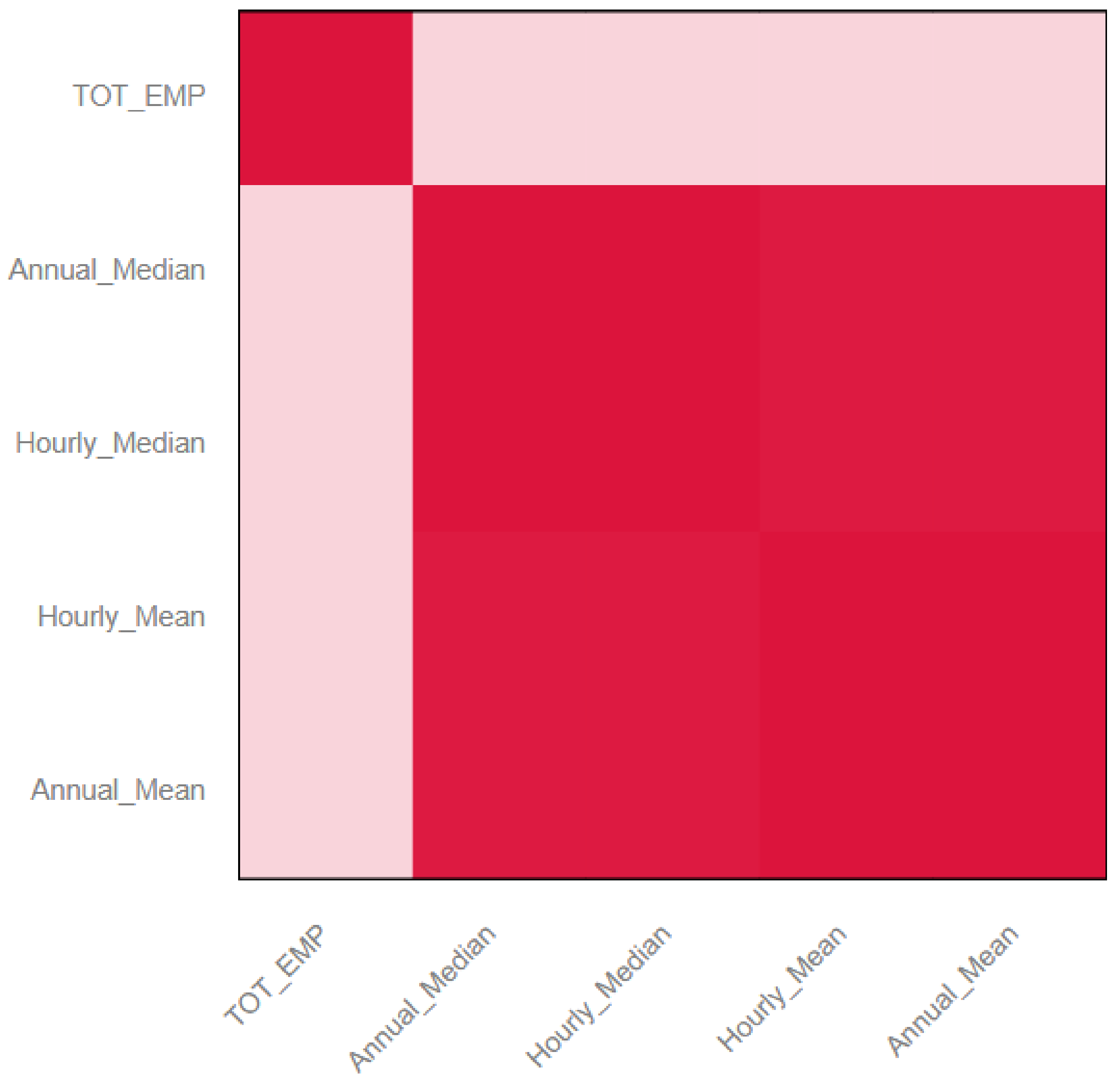

TOT_EMP
Annual_Median
Hourly_Median
Hourly_Mean
Annual_Mean
TOT_EMP
Annual_Median
Hourly_Median
Hourly_Mean
Annual_Mean

**Fig:17**

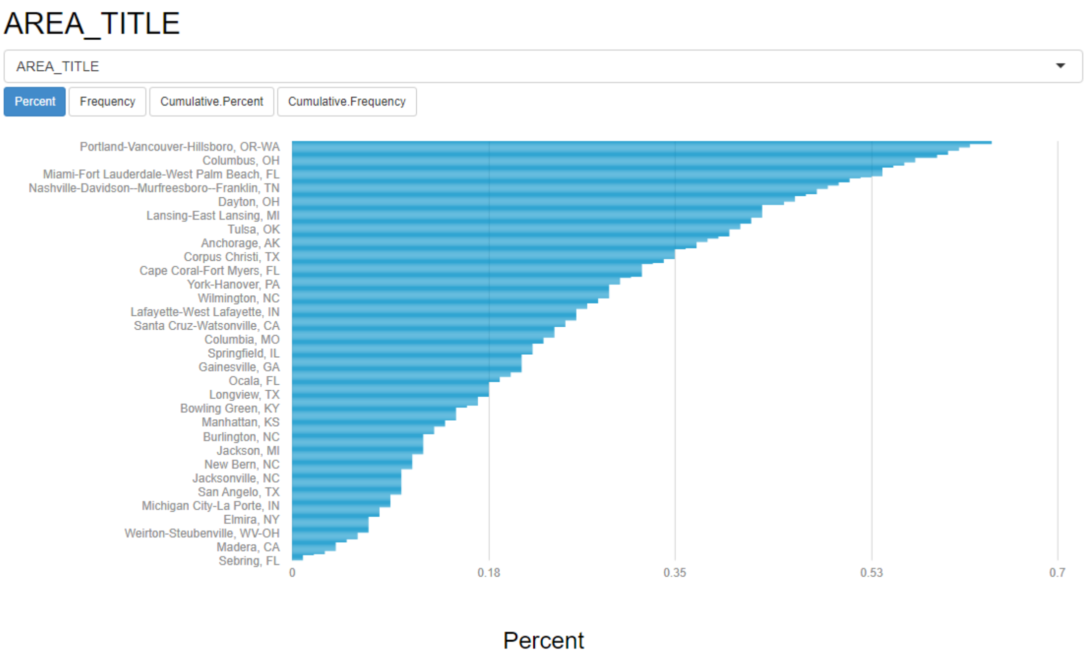

AREA_TITLE
AREA_TITLE
Percent
Frequency
Cumulative.Percent
Cumulative.Frequency
Portland-Vancouver-Hillsboro, OR-WA
Columbus, OH
Miami-Fort Lauderdale-West Palm Beach, FL
Nashville-Davidson--Murfreesboro--Franklin, TN
Dayton, OH
Lansing-East Lansing, MI
Tulsa, OK
Anchorage, AK
Corpus Christi, TX
Cape Coral-Fort Myers, FL
York-Hanover, PA
Wilmington, NC
Lafayette-West Lafayette, IN
Santa Cruz-Watsonville, CA
Columbia, MO
Springfield, IL
Gainesville, GA
Ocala, FL
Longview, TX
Bowling Green, KY
Manhattan, KS
Burlington, NC
Jackson, MI
New Bern, NC
Jacksonville, NC
San Angelo, TX
Michigan City-La Porte, IN
Elmira, NY
Weirton-Steubenville, WV-OH
Madera, CA
Sebring, FL
0
0.18
0.35
0.53
0.7
Percent

**Fig:18**

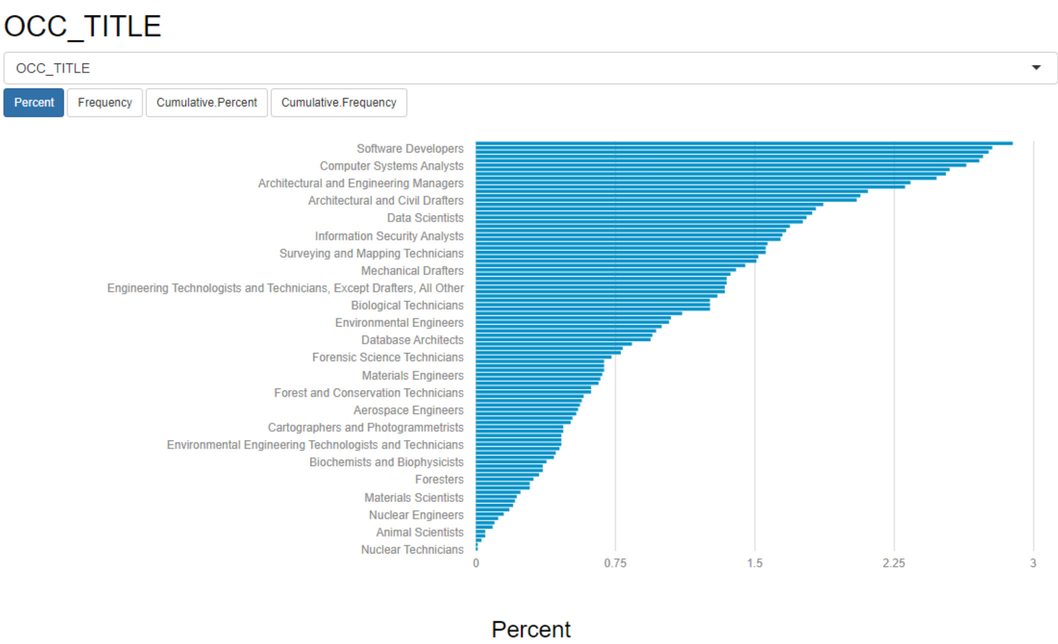

OCC_TITLE
OCC_TITLE
Percent
Frequency
Cumulative.Percent
Cumulative.Frequency
Software Developers
Computer Systems Analysts
Architectural and Engineering Managers
Architectural and Civil Drafters
Data Scientists
Information Security Analysts
Surveying and Mapping Technicians
Mechanical Drafters
Engineering Technologists and Technicians, Except Drafters, All Other
Biological Technicians
Environmental Engineers
Database Architects
Forensic Science Technicians
Materials Engineers
Forest and Conservation Technicians
Aerospace Engineers
Cartographers and Photogrammetrists
Environmental Engineering Technologists and Technicians
Biochemists and Biophysicists
Foresters
Materials Scientists
Nuclear Engineers
Animal Scientists
Nuclear Technicians
0
0.75
1.5
2.25
3
Percent

**Fig:19**

## Model Comparison Report

### Fit and error measures

| Model | Correlation | RMSE | MAE | MPE | MAPE |
|---|---|---|---|---|---|
| Forest_Model | 0.5734 | 28,075.6003 | 20,304.1891 | -4.9450 | 22.4089 |
| Decision_Tree_ | 0.5733 | 25,363.9830 | 17,674.6983 | -3.0012 | 18.4226 |
| Boosted_Model | 0.5888 | 25,062.0593 | 17,200.6840 | -2.5139 | 17.7364 |

Model: model names in the current comparison.
Correlation: correlation between the predicted values and the actual values.
RMSE: root mean square error.
MAE: mean absolute error.
MPE: mean percentage error. Note: based on its definition, MPE may be positive or negative infinity if the target variable has 0 values. In this case, we return a weighted percentage error (WPE).
MAPE: mean absolute percentage error. Note: based on its definition, MAPE may be positive or negative infinity if the target variable has 0 values. In this case, we return a weighted absolute percentage error (WAPE).

### Performance Diagnostic Plots

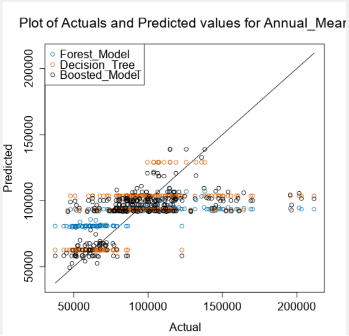
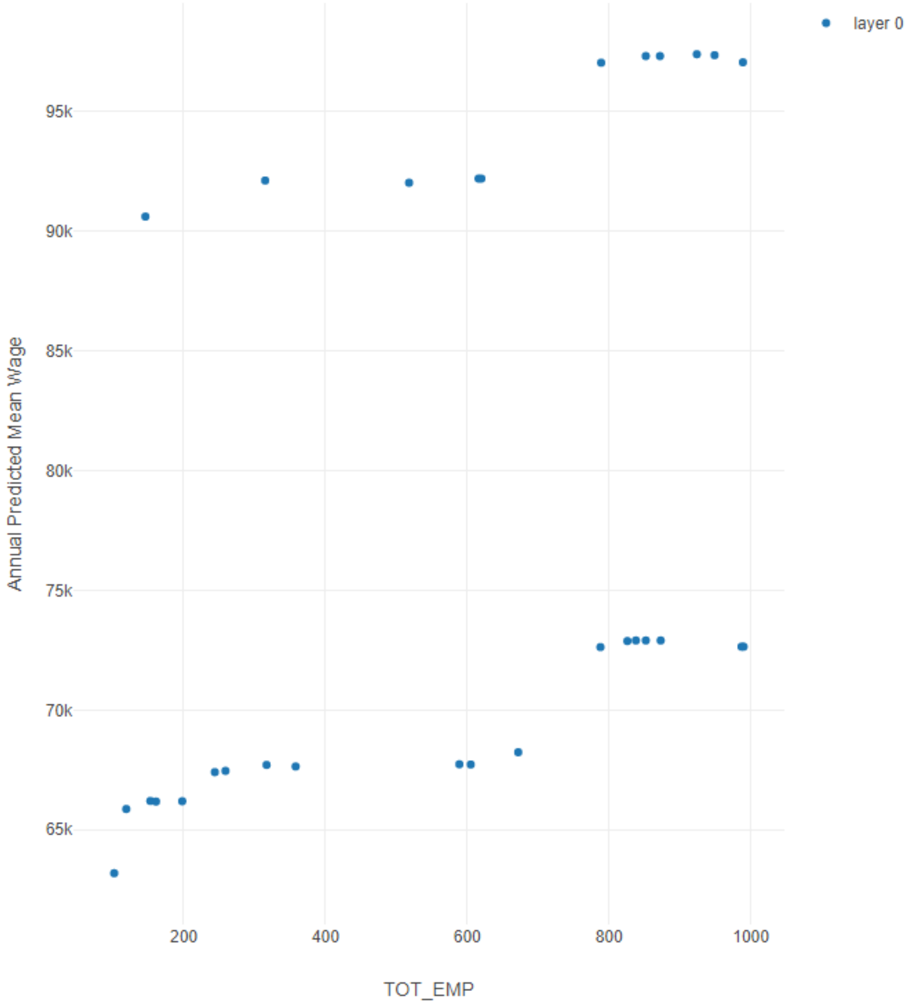



**Fig:20**



layer 0
95k
90k
85k
80k
75k
70k
65k
Annual Predicted Mean Wage
200
400
600
800
1000
TOT_EMP


**Fig:21**