

## **Data 608 Story 4**

**By Jean Jimenez**

### **Exploring The Salary of a Data Practitioner in the United States**

#### **Abstract:**

This study explores the distribution of salaries for data practitioners across different geographic locations, job roles, and industries in the United States, leveraging data from the Bureau of Labor Statistics' Occupational Employment and Wage Statistics (OEWS) for May 2022. By analyzing the most recent employment data, the research aims to provide comprehensive insights to guide data science professionals in making informed career decisions. The methodology involved extracting annual wage estimates for approximately 830 occupations, focusing on six data practitioner roles identified via the Standard Occupational Classification (SOC) system. The analysis utilized Excel for data visualization and R for data processing, examining metrics such as annual salary by state, SOC code, and industry type.

Results indicate significant regional variations in mean and median salaries, with the highest salaries observed in the northeastern United States, particularly in Connecticut, Delaware, and Maryland. Data Scientist roles garnered the highest mean annual salary, while the Information, Communication, and Technology industry emerged as the top-paying sector. The study reveals disparities in salary distributions, highlighting the influence of geographic location, job title, and industry on compensation.

The discussion section delves into the economic and demographic factors contributing to these disparities, suggesting the median salary as a more reliable metric for evaluating compensation due to the potential skewing effect of outliers. It also notes the high variance in salaries within certain states and industries, suggesting further research into emerging markets and the impact of industry-specific demands on salary trends.

In conclusion, the study underscores the importance of geographical location, job title, and industry in determining salary levels for data practitioners. It suggests that entry-level professionals may find attractive opportunities in emerging markets, despite the allure of higher salaries in established regions and sectors. Further research incorporating historical data could offer deeper insights into long-term salary trends and their implications for career planning in the data science field.

#### **Introduction:**

As a candidate pursuing a Master of Science in Data Science, understanding the landscape of salary distributions across various geographic locations and job roles is crucial. This insight is not only instrumental in informing our employment search strategies but also in selecting the most promising industries and roles to explore. Through an analysis of the latest employment data, I aim to equip individuals with detailed information to make well-informed career decisions.

## **Methods:**

The data used for this analysis was from the [Bureau of Labor Statistics' Occupational Employment and Wage Statistics \(OEWS\)](#). The BLS OEWS is a data set with annual wage estimates for ~830 occupations divided by different levels. The data was accessed on 03/11/2024. The most recent version of the data was used (May 2022 publication). Raw data was downloaded in .xls format.

The BLS standardizes occupational titles using the Standard Occupational Classification (SOC). Each occupation has a unique SOC code. Data practitioner roles were entered into [O-Net](#) and 6 related SOC codes were returned (corresponding to Data Scientist, Statistician, Operation Research Analyst, Mathematical Science Occupations, Biological Scientists, and Financial Specialists).

Metrics collected were annual salary by state, SOC code, and industry type. Excel was used to analyze the data, as well as make the figures. R was used to assist with the processing of data, as well as the generation of tables.

## **Results:**

### ***By State***

Connecticut has the highest mean salary (\$110.76 k) followed by Delaware (\$103.64 k), Maryland (\$100.92 k), New York (\$100.64 k), and California (\$100.19 k). The states with the lowest mean salary for data practitioners are Missouri (\$72.16 k), Kansas (\$70.93 k), Arkansas (\$70.43 k), West Virginia (\$69.47 k), and Louisiana (\$67.46 k) (figure 1).

Data practitioners had the highest median salary in Delaware (\$114.92 k), followed by Connecticut (\$105.81k), Maryland (\$104.26 k), New Hampshire (\$102.87 k), New Mexico (\$100.75 k). The states where data practitioners had the lowest median salary are Montana (\$72.10 k), Florida (\$71.43 k), Wisconsin (\$70.12 k), Louisiana (\$67.43 k), and Kansas (\$66.87 k) (figure 2).

The states with the most variance (measured using standard deviation) are Wyoming (+/- \$30.95 k), followed by California (+/- \$26.19 k), Washington (+/- \$22.20 k), North Carolina (+/- \$20.87 k), and New Jersey (+/- \$20.00 k). The states with the least variance in annual salary are Arkansas (+/- \$8.15 k), followed by Nebraska (+/- \$7.88 k), West Virginia (+/- \$7.15 k), Vermont (+/- \$5.67 k), and Montana (+/- \$4.07 k) (figure 3).

### ***By Job Title***

The data scientist role experienced the highest mean annual salary (\$95.95 k +/- \$19.16 k). This is followed by statisticians (\$93.20 k +/- \$21.48k), Operations Research Analysts (\$85.42 k +/- \$15.21 k), Mathematical Science Occupations (\$84.76 k +/- \$22.84 k), Biological Scientists (\$81.12 k +/- \$14.53 k), and Financial Specialists (\$74.20 k +/- \$14.75 k) (Table 1)(figure 4).

### ***By Industry Category***

The industry category with the highest mean salary for data practitioners is Information, Communication, and Technology (\$109.57k +/- \$18.10k), followed by Professional, Scientific, and Technical Services (\$105.70k +/- \$14.84k), Manufacturing and Construction (\$102.70k +/- \$27.09k), Retail and Wholesale Trade (\$101.21k +/- \$24.61k), Financial Services, Insurance, and Real Estate (\$96.94k +/- \$16.81k), Other Industries (\$89.80k +/- \$21.37k), Healthcare and Social Assistance (\$84.32k +/- \$16.01k), and finally Accommodation, Food Services, and Entertainment (\$81.58k +/- \$16.98k)(Table 2)(Figure 5).

### **Discussion:**

The greater northeastern mega-region is the area of the United States where data practitioners have the highest mean salary (with an exception of California). This is consistent with that region being the most populous region in the United States. This region has more jobs that pay on average more than other states (when compared to the lowest average paying states). The lowest paying states are consistent with states with extreme poverty/low GDP (like West Virginia) or extremely low population (like Montana). California also having a high mean salary is probably consistent with the tech industry and driven by the high wages these company pays their employees.

When it comes to looking at salary earned for data practitioners, mean annual salary might be a bit of a misleading metric to use. The distribution of salaries might shift the mean away from the median and make it an inaccurate measure (especially if there are a lot of outliers). Sometimes, median is a more appropriate metric to use. For states where the mean annual salary was greater than the median, there are a smaller number of high earners. This was true for New York. Out of the 5 states with the highest average salary, 2 States (NY and CT) had lower median annual salary (although California was not much of a difference compared to NY).

States that have a higher median salary when compared to mean annual salary might not have enough data practitioner jobs/ data on this type of salary. It might also mean that these states might have a higher percentage of entry-level lower-paid data practitioners

(which would drag the mean down). New Mexico and New Hampshire are consistent with this. This would make sense for New Mexico, since there have been recent investments being made in the precious metal mining and clean energy industries (possibly creating more entry level jobs).

Standard deviation was used as a metric to account for the variance in average annual salary between the states. It would be ideal to work in a state with the highest mean annual salary and lowest standard deviation. A lower standard deviation would signify that the distribution of salary is closer to that mean (meaning you have a greater chance of finding similar-paying roles). A higher standard deviation would mean great spread in the salary. California, while having a high mean annual salary, also has a high standard deviation. In CA, you could be making ~\$30k less than mean (depending on the job title or industry).

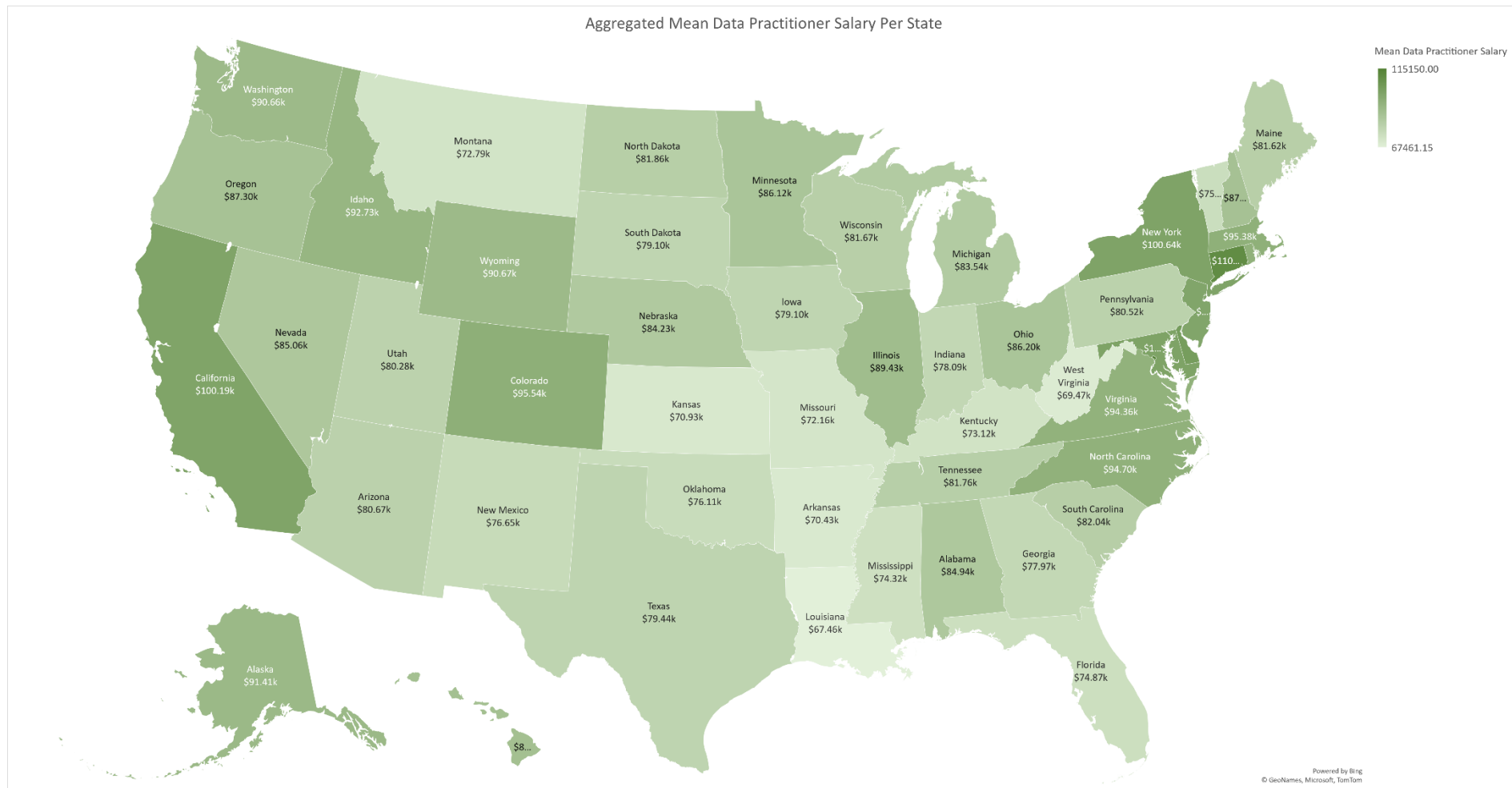
When it came to looking at salary by job title, the data scientist role made the highest annual mean salary. Lowest paid were data practitioners in biological sciences and financial roles. When it came to looking at salary by industry type, Communication/ IT, followed by Professional /scientific/ technical paid the most on average. This makes sense with the record profits tech and biotech companies make. Lowest paid industries are healthcare/social assistance, and entertainment/ accommodation /food services.

### **Conclusion:**

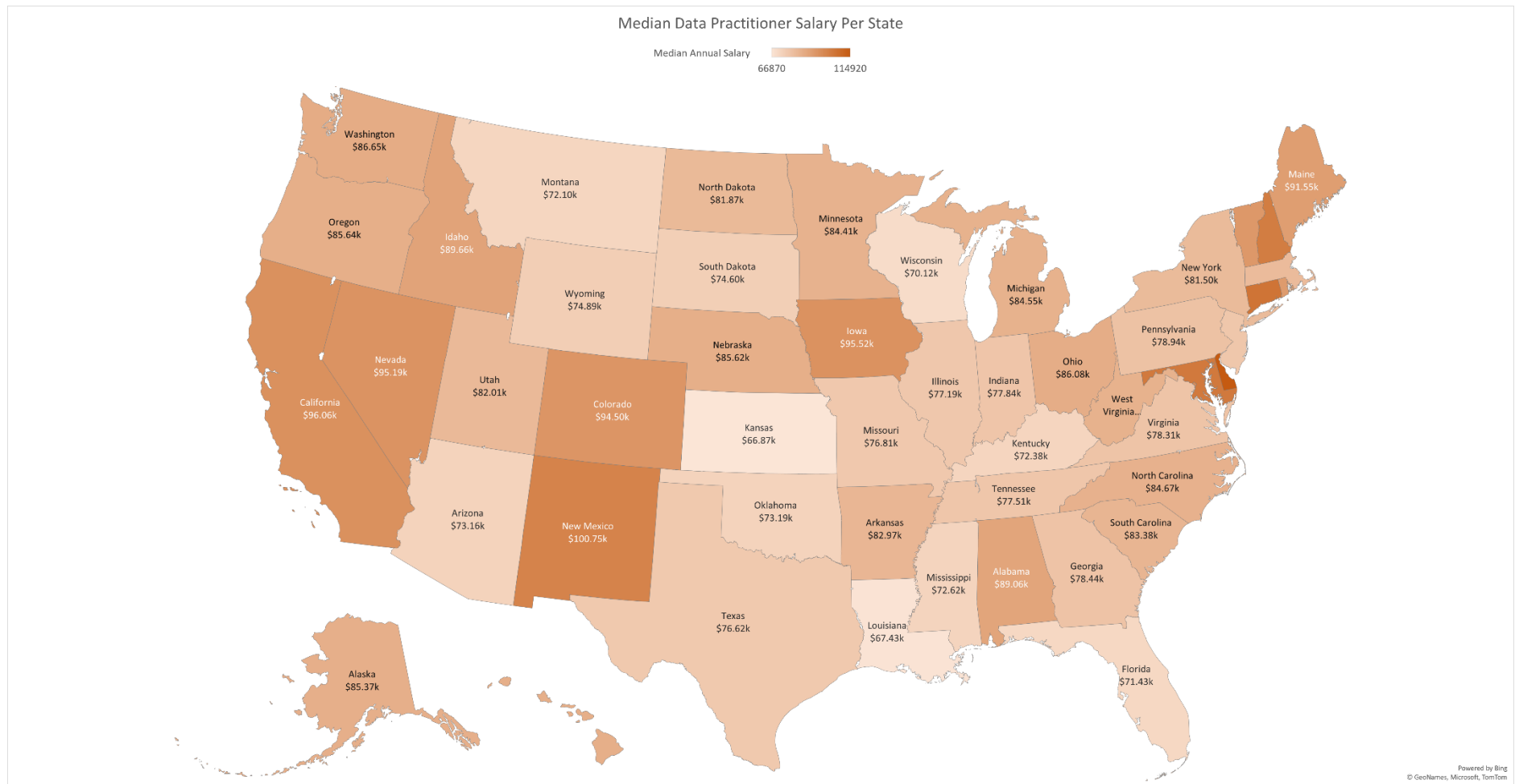
The north-eastern to mid-Atlantic region of the United States are the areas where data practitioners make on average the most (in terms of annual salary). More analysis needs to be done on states with higher median salary. These states might have new emergent markets and might be favorable and attractive to entry-level data practitioners (even though they might be attracted to the high salary of the other states). It might be beneficial to incorporate historical data into the model to analyze the trends per state over time.

Data scientist role got paid on average the most, while biological and financial data practitioners got paid the least. Communication/ IT, followed by Professional /scientific/ technical paid the most on average. Lowest paid industries are healthcare/social assistance, and entertainment/ accommodation /food services. More work should be done exploring why the healthcare sector is the lowest paid (but arguably the most important).

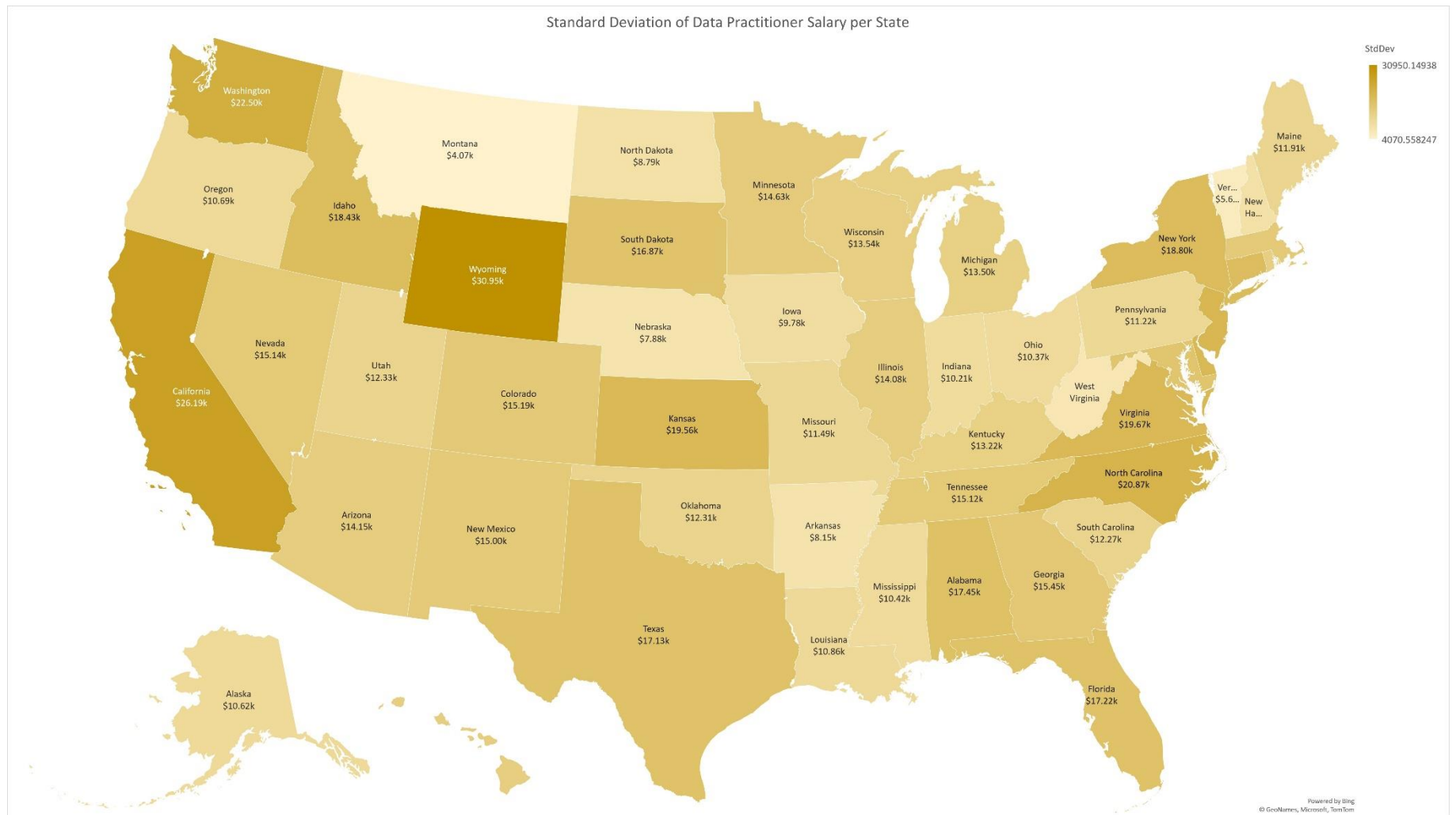
### **Figures and Tables:**



**Figure 1-** Mean Annual Salary by State for Data Practitioners (\$ USD)



**Figure 2-** Median Annual Salary by State for Data Practitioners (\$ USD)

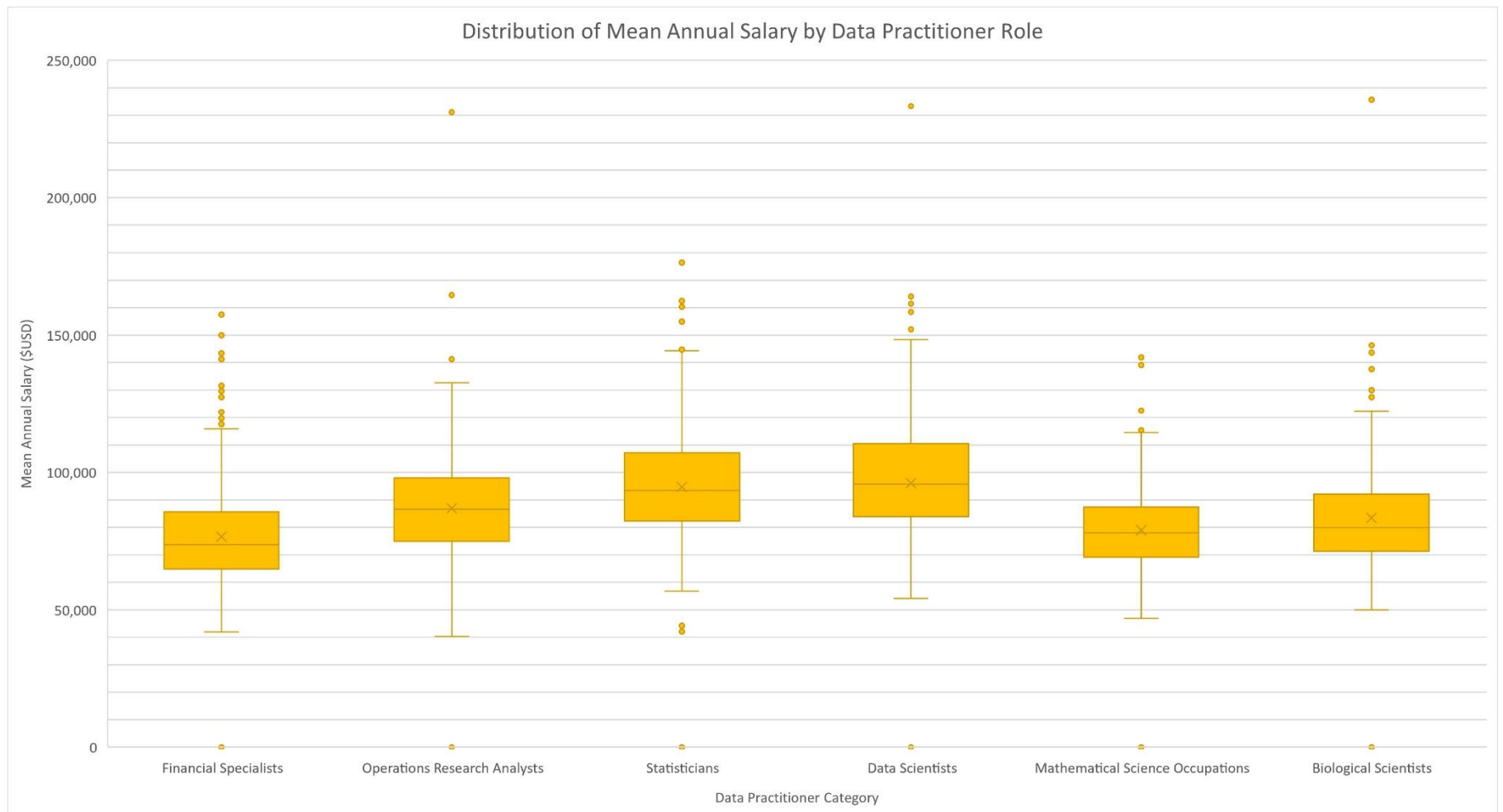


**Figure 3-** Standard Deviation of Annual Salary per State of Data Practitioners

**Table 1-** Statistics of Annual Salary by Occupation Title

OCC_TITLE	Min	Q1	Median	Mean	StDev	Q3	Max	IQR
Data Scientists	54120	84770	96260	98459.89	19490.73	110912.5	233320	26142.5
Statisticians	42020	82580	93620	95864.04	21358.24	107360.0	176420	24780.0
Operations Research Analysts	40350	75345	86710	87677.43	18361.00	98085.0	231150	22740.0
Biological Scientists	49990	71575	80030	84240.16	20963.77	92120.0	235650	20545.0
Mathematical Science Occupations	46850	69660	78220	81055.33	17695.42	87385.0	141860	17725.0
Financial Specialists	41900	65050	73740	77116.89	18303.09	85657.5	157410	20607.5

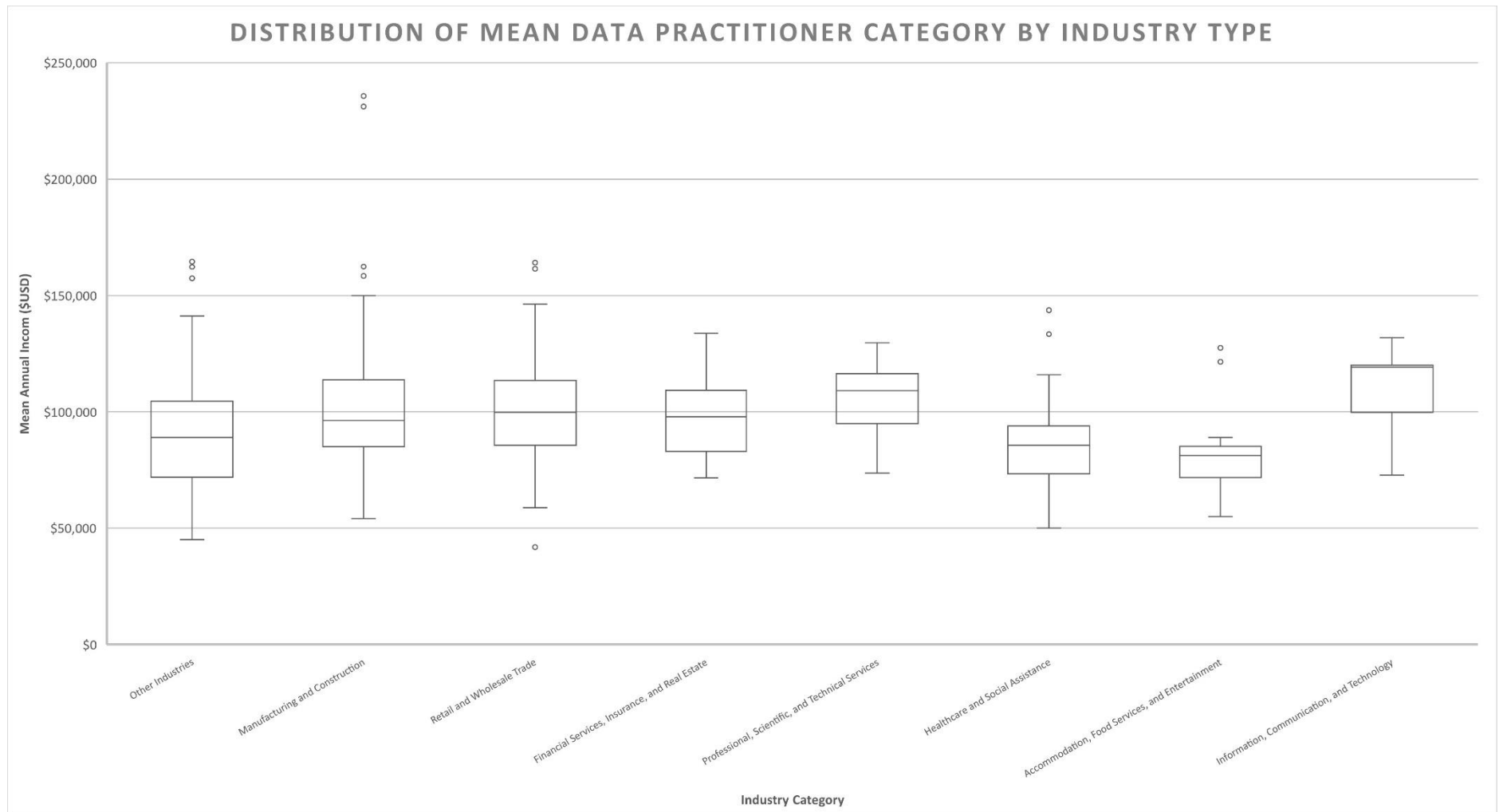




**Figure 4-** Distribution of Annual Salary by Occupation Title

**Table 2-** Statistics of Annual Salary by Industry Category

category	Min	Q1	Median	Mean	StDev	Q3	Max
Information, Communication, and Technology	72850	99770.0	119170	109570.40	18101.05	120100	131840
Professional, Scientific, and Technical Services	73640	95190.0	109160	105695.86	14841.10	115850	129630
Manufacturing and Construction	54120	85200.0	96290	102704.38	27093.43	113740	235650
Retail and Wholesale Trade	41900	85660.0	99780	101209.83	24605.96	113160	164030
Financial Services, Insurance, and Real Estate	71540	83147.5	97830	96936.43	16813.59	108050	133730
Other Industries	45110	71910.0	89005	89796.30	21368.10	104355	164550
Healthcare and Social Assistance	50080	73470.0	85600	84324.86	16008.55	93980	143680
Accommodation, Food Services, and Entertainment	55040	72030.0	81260	81576.90	16977.88	85180	127420



**Figure 5-** Distribution of Annual Salary by Industry Category