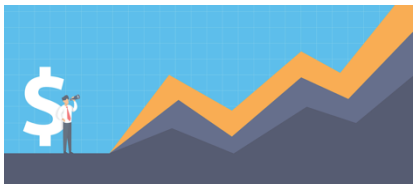
SALARY

By Quoc Hung Nguyen, Lily Slater, Louis Perruchon, Victor Loyola, Xuefei Li



DATA-RELATED JOBS, LOCATIONS AND COMPANIES

WITH USE OF THE visa-related SALARY DATA SET FROM THE US DEPARTMENT OF FOREIGN LABOR certificate
http://www.foreignlaborcert.doleta.gov/performancedata.cfm

Objectives

- Compare salaries among different sub-types of data-related jobs to see which job has higher or lower salaries than others.
- Compare salaries of data-related jobs among different states to see which states have the highest paying data-related salaries.
- Compare offered salaries to the prevailing wage to see which jobs get overpaid or underpaid.

JOB TITLES

LOCATIONS

SALARIES

STANDARD OF LIVING

Executive Summary

Description of the Report

For this report, our group analyzed the salary data set, aiming to assemble and present key indicators of the workforce in data-related jobs in the United States for the students of DS311. It is important for students to understand the workforce in the area of work they are interested in because it can help them make more informed decisions about their future career. Knowing about the size of the workforce, the types of jobs available and the salaries that can be earned can help students determine how competitive the job market is and as well as which job titles to pursue and what salary range to expect. Having this knowledge can help students make more informed decisions about their career path and set realistic expectations for their future.

Description of the Data

A US corporation must apply to the US government to obtain a green card or visa for their foreign candidates when they hire someone from outside the country for a technical role. With the help of these programs, the US government can keep tabs on whoever is traveling for business purposes and who is leaving the country for personal reasons. To ensure that immigrants are not being exploited, companies must disclose the employee's anticipated wage each time they file a visa or green card application to ensure wage equality for them and their US citizen counterparts. Additionally, they must disclose the normal salary for a worker with comparable qualifications.

The original data was compiled by the US Department of Labor's Office of Foreign Labor Certification. This data includes 167,278 observations for the case status of permanent resident applications from 2008 to 2015. There also have 27 columns that represent 27 variables that statistically list: Case number, Case status, Case Received Date, Decision date, Employer Name, Prevailing Wage Submitted, Prevailing Wage Submitted, Unit, Paid Wage Submitted, Paid Wage Submitted Unit, Job Title, Education Level Required, College Major Required, Experience Required Y/N, Experience Required Num Months, Country of Citizenship, Prevailing Wage Soc Code, Prevailing Wage Soc Title, Work City, Work State, Work Postal Code, Full Time Position Y N, Visa Class, Prevailing Wage Per Year, Paid Wage Per Year and Job Title Subgroup.

I - Data-related Jobs Compared to Other Types of Jobs

Introduction:
We made the choice to focus on data analyst and data scientist jobs as data related jobs. We didn't take in consideration Software Engineer and Business analyst, even though they can touch to data. We considered that this not their main task to deal with it. Data Analyst and Data scientist are doing different tasks around data, but their common point is that they are always dealing with data questions. Consequently, we took data analyst and data scientist as data related jobs.

A- Do specific Sub-types of data related jobs have higher or lower salaries than others?
For data scientist we can see in average that they do around 110K per year and data analyst 70K per year. The maximum amount for these two categories it can growth until 230 k for a data scientist and 300K for a data analyst.

If we take a compared them to others, we can say that they are common. Not the best salaries not the lowest. The best is attorney who has in average a wage of 145K, and the lowest is teacher with an average wage of 47K per year.

B- Which companies have the highest salaries for those sub-types?
If we look to the compagnies with the highest salaries for data related jobs we can see that for a data analyst, the corporation "Intuit" is in first position. For data scientist it's Netflix who hold the first position.

We can take that information as insight for our next question, by looking at specificities. Netflix has its headquarters in California, in the south of the Silicon Valley at Los Gatos. Intuit has also her headquarter and majority of their workspaces in California.

C- Changes with location of the job

For Data related jobs, we can observe as we seen above with the maximum amount, that that California is the leader. In fact, for data scientist jobs, California holds the first position with 117k in average per year. For Data Analyst, California is second with 80 470 $ per year, just after Connecticut with 80 989$ in average. SO, they are a approximately a difference of 500 dollars between the two states which is marginal on a year (less than a hundred dollar per month, for salaries around 7k per month, so less than 2% of difference).

D- Standard of living into account.

We can acknowledge by the US Bureau of economic Analysis, that the nominal GDP per habitant is one of the highest from the US states range, around 80k a year.

For the cost of living of this state, it's also the maximum of USA with an index of 139.7. Despite this fact if we look closely how the cost of living is made, we can see that it takes in account lot of marginal variable compared to the huge amount of money made by a data related jobs. Consequently, a data analyst or a data scientist should go to California.

Conclusion:

We can conclude for this part that, a data Analyst or a data scientist should move to California, because it's the place where he can have the highest salaries even though the standard of living is expensive.

From a less objective point of view, we can add that because majority of the tech company are in California with its Silicon Valley, they are a lot of opportunities for data related jobs. Consequently, the cost of living can be seen as only as investment on yourself, due to a lot of opportunities present in California.

II - States with the highest Salaries for data-related jobs

What states (of those I am willing to move to) have the highest paying data-related salaries?

To present this information, we wanted to see the top 5 states for both positions. We thought if you were using this information to pick a state to move to for work, you would want a location that has a high salary for both positions in case perhaps you start as an analyst then move to being a data scientist. The 5 states that scored highest when averaging their salaries for these positions and they were New York, Minnesota, Connecticut, Washington and California. The average salaries for these states, for each data related job, are displayed in the plot (see graph 2). California is typically the highest paying state for data analysts and data scientists, with average salaries ranging from $80,000 to over $120,000. Washington, Connecticut, and Minnesota also have high salaries, averaging around $70,000 to $110,000. New York is also a high-paying state, with salaries averaging around $70,000 to over $110,000.

Differences between job sub-categories?

Data analysts and data scientists both work with data and develop insights from it, but they have different roles and responsibilities. You can see in the graph that salaries for data scientists tend to higher than data analyst positions. In general, data analysts have more experience with data analysis and reporting, while data scientists have more experience with machine learning and predictive analytics. As a result, data scientists typically get paid more than data analysts due to their more complex and varied roles.

Which companies have the highest salaries for those sub-types?

To answer this question, we found the companies that paid the highest salary for these positions in each of these five states (see table 2). For California, netflix paid the highest salaries for both data analyst ($220,000) and data scientist ($230,000) roles. Other well-known companies that appear include JPmorgan, Microsoft, Target and GE. These larger companies tend to have the most advanced data sets, require the most sophisticated solutions, and offer the best career advancement opportunities.These salaries are not actually indicative of the highest salaries in California which can reach higher rates than $3000,000. This is a reminder that this visa-related dataset is not representative of the actual range of salaries in these positions in the United States.

III - Comparing offered salaries to the prevailing Wage

Are there job-subcategories that tend to get over-paid or under-paid?

By definition, overpaid means that the actual salary per year of an employee is greater than the prevailing salary per year the employee expected to have while underpaid means that the prevailing salary per year of an employee is greater than the salary they received in reality.

For job sub-categories, there is a large difference between for the mean received and the mean prevailing for assistant professor and attorney, which means that these two job categories seem to get overpaid. These values for business analyst, data analyst are similar. However, for more details, 77% and 70% of data analyst and business analyst employees get overpaid, respectively, which means that the wage difference between the prevailing and the reality should be not significant. In contrast, teacher, software engineer and data scientist is slightly overpaid, while management consultant employee is about to get the salary of what they expected.

Are there companies that tend to get over-paid or under-paid?

For companies, the percentage of getting overpaid and underpaid is only counted for companies with more than or equal to 100 employees. Since the data is from the visa applications, which means that there are many Native American workers who still work in the company and do not need to apply for a visa. Therefore, if the companies does not have many employees in this dataset, the conclusion that the company is getting overpaid/underpaid is not significant.

Overall, small company tend to be overpaid while large companies such as Fujitsu America and Cisco system tend to be underpaid. For more details, Argha Services INC, Astir It Solutions INC, Cedent Consult INC, Fort Worth Independent School district, Pegasus Infotech INC, Primesoft INC, Promatrix Corp, SCM Data INC, Sri Anjaneya Technologies, LLC and Vabsys INC are companies with more than 100 employees that have 100% rate of overpaid, while Cognizant Technology Solutions US Corporation, Fidelity Brokerage Services LLC, Fidelity Technology Group LLC, HCL America, Igate Technologies INC, Indsoft INC, Los Angeles Unified School District and New York City Department of Education are companies with more than 100 employees that have 100% rate of underpaid.

Will the answer change if I take standard of living into account?

The standard cost living into account of each state in America does seem to affect the overpaid/underpaid rate since the people living in high standard cost living, which is Oregon, Washington, Alaska, and California, get underpaid. The reason might be because that since they live in an expensive place, they want to have higher salary to afford their living cost.

IV - Additional Questions of the Dataset

Because California kept coming up in our results and because we all currently live here, we chose to focus on California for our additional questions. California is a great place to live and work in data because it offers a large and diverse economy and a wealth of opportunities. It is home to some of the world's leading technology companies, such as Google, Apple, and Facebook and excellent universities and research institutions, which are hubs for data science and related fields. Finally, California is an attractive destination for data professionals because of its sunny weather, vibrant culture, and exceptional quality of life.

In California,which cities have the highest salaries?

We looked at cities in California with the highest salaries (plot 9). The green bars in the plot indicate data scientists and blue is data analyst. And is it filtered to cities that had average salaries of 120,000 or more. Only three cities in California had data analyst salary averages of 120,000 and over. There were Redwood Shores, Menlo Park and Los Gatos. There were more cities that had average salaries of 120,000 and over for data scientist positions. They included, San Jose, Brisbane, San Ramon, San Bruno, Pleasanton, Carlsbad, Cupertino, Menlo Park and Los Gatos. This representation shows us again that you are likely to earn more in a data science position rather than data analytics. It also tells us that there are high salaries available in Menlo Park and Los Gatos for both of these positions. This isn't surprising as Menlo Park is located in the heart of Silicon Valley, which is home to some of the most innovative companies in the world. This means that there are ample opportunities to find employment in a wide range of data-related roles and industries. Los Gatos is also close to silicon valley and home to a number of companies specializing in data science and analytics, as well as many other tech companies. Additionally, the cost of living in Los Gatos is quite reasonable, making it a great place to live and work.

In California, which cities have the highest overpay/underpay rate?

We looked at the cities in California that have the most overpaid/underpaid rate. From the graph, it seems that cities from the bay area get overpaid, except for the cities with the density of people is bigger, which is San Jose and San Francisco. Similar explanation with the standard cost living into account, big city seems to have more activities, therefore, they tend to pay more compare to those other cities near by. As the result, they expected their salary should be more which means that underpaid rate in these cities is higher.

The day of processing does affect the result of the case application?

From the table, if the time duration for a case status is short, the case status is tended to get approved, while if it takes more time, case status is tended to get denied.

V Technical Appendix

Graph 1:

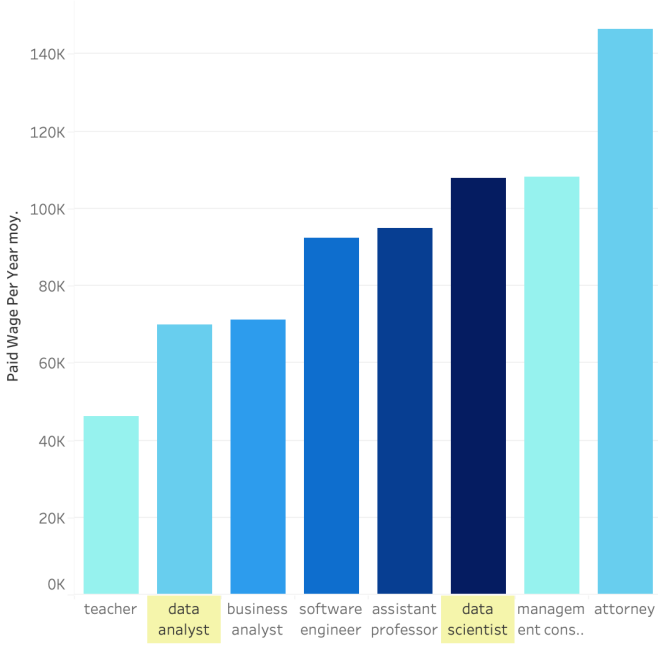
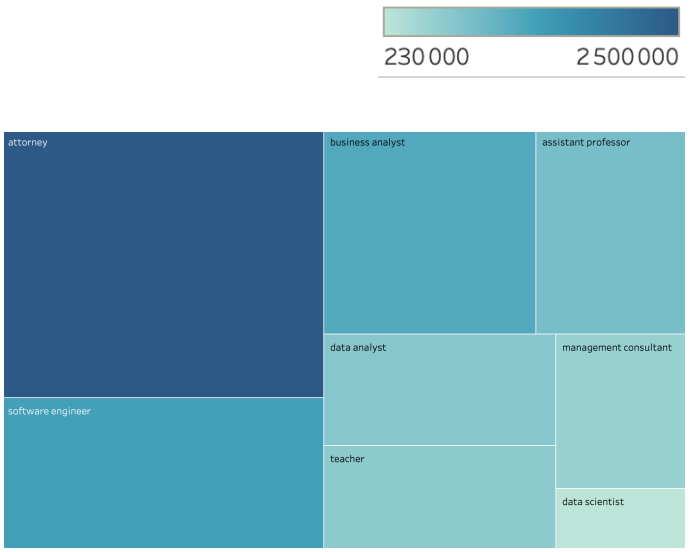


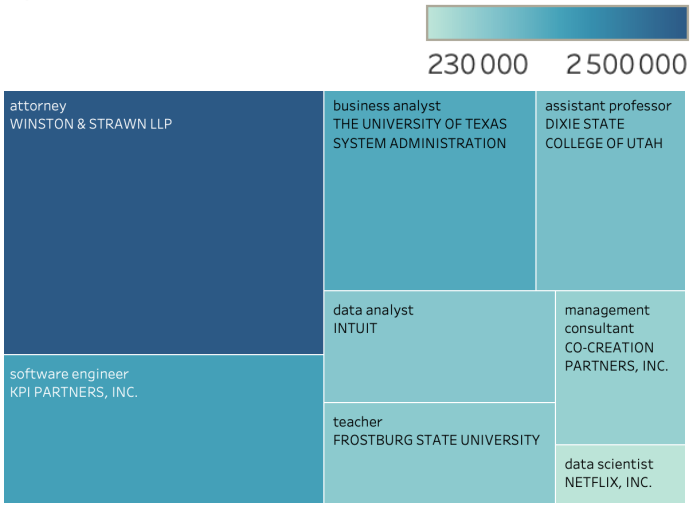
Table 1:

data analyst	Maine	72 595
	New Hampshire	75 943
	Delaware	78 561
	California	80 470
	Connecticut	80 989
data scientist	Oregon	106 141
	New York	106 332
	Minnesota	111 457
	Washington	116 133
	California	117 824

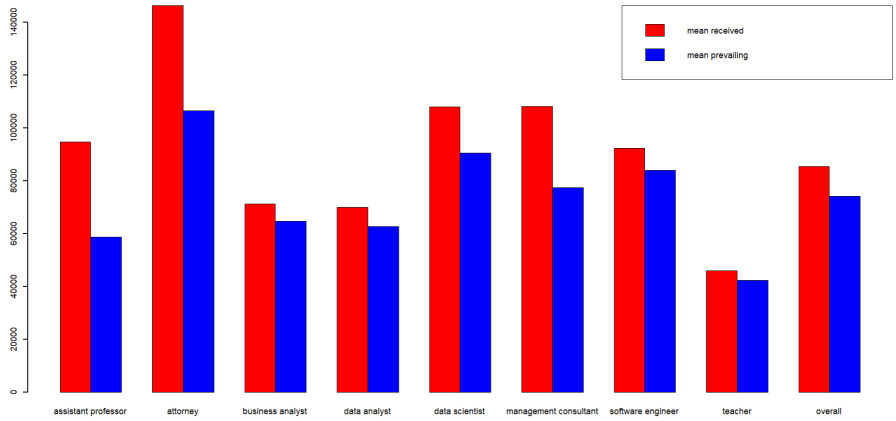
Plot 1:



Plot 2:



Graph 3:



Graph 4: Underpaid/overpaid rate for each job-sub categories

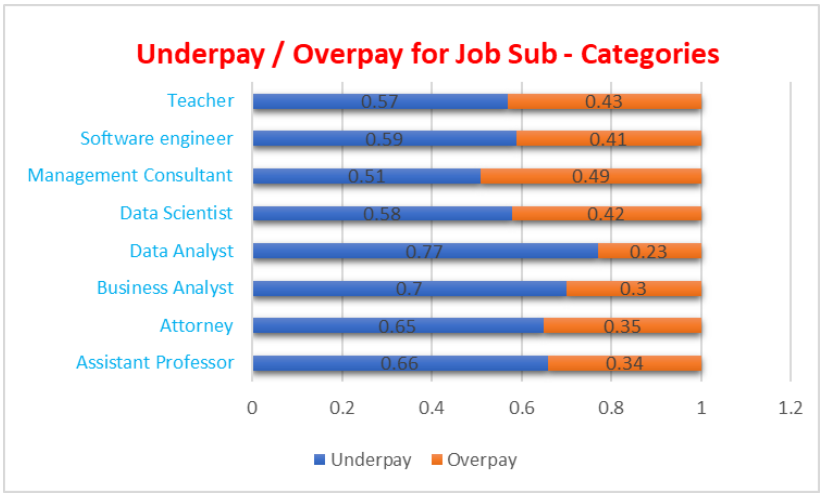


Table 3: Underpaid Companies

	EMPLOYER_NAME	Number of Employees	Overpay	Underpay	Underpay rate
1	COGNIZANT TECHNOLOGY SOLUTIONS US CORPORATION	141	0	141	1
2	FIDELITY BROKERAGE SERVICES LLC	102	0	102	1
3	FIDELITY TECHNOLOGY GROUP LLC	227	0	227	1
4	HCL AMERICA, INC.	134	0	134	1
5	IGATE TECHNOLOGIES, INC.	151	0	151	1
6	INDSOFT, INC	112	0	112	1
7	LOS ANGELES UNIFIED SCHOOL DISTRICT	167	0	167	1
8	NEW YORK CITY DEPARTMENT OF EDUCATION	311	0	311	1
9	BALTIMORE CITY PUBLIC SCHOOLS	411	3	408	0.9927007
10	CSC COVANSYS CORPORATION	183	2	181	0.989071
11	FUJITSU AMERICA, INC.	1814	20	1794	0.9889746
12	INTEL MOBILE COMMUNICATIONS NORTH AMERICA	130	2	128	0.9846154
13	NATIONAL FINANCIAL LLC	106	2	104	0.9811321
14	DENVER PUBLIC SCHOOLS DISTRICT 1	101	2	99	0.980198
15	CISCO SYSTEMS, INC.	1011	27	984	0.9732938

Table 4: Overpaid companies

	EMPLOYER_NAME	Number of employees	Over pay	Under pay	Overpay rate
1	ARGHA SERVICES, INC	260	260	0	1
2	ASTIR IT SOLUTIONS INC.	111	111	0	1
3	CEDENT CONSULTING INC	153	153	0	1
4	FORT WORTH INDEPENDENT SCHOOL DISTRICT	111	111	0	1
5	PEGASUS INFOTECH, INC.	127	127	0	1
6	PRIMESOFT, INC	122	122	0	1
7	PROMATRIX CORP	150	149	0	1
8	SCM DATA, INC	284	284	0	1
9	SRI ANJANEYA TECHNOLOGIES, LLC	108	108	0	1
10	VABSYS, INC.	100	100	0	1
11	DALLAS INDEPENDENT SCHOOL DISTRICT	808	807	1	0.998762
12	INTRAEDGE, INC.	331	330	1	0.996979
13	VISION IT SERVICES USA INC	243	242	1	0.995885
14	MMC SYSTEMS INC.	176	175	1	0.994318
15	V-SOFT CONSULTING GROUP, INC	129	128	1	0.992248

Table 5+6: Overpaid/Underpaid rate for States in America

	WORK STATE	count	overpay	underpay	overpay rate	Cost of living rank
1	Puerto Rico	109	101	8	0.9266055	29
2	North Dakota	138	124	14	0.8985507	23
3	Wyoming	45	39	5	0.8863636	14
4	Virginia	109	95	14	0.8715596	32
5	South Dakota	105	90	15	0.8571429	23

	WORK STATE	count	overpay	underpay	underpay rate	Cost of living rank
1	Oregon	1535	581	951	0.6207572	45
2	Washington	4610	2096	2511	0.5450402	40
3	Alaska	66	34	32	0.4848485	47
4	New Mexico	575	298	277	0.4817391	20
5	California	46782	24379	22391	0.4787471	49

Table 8+9: Overpaid/Underpaid for Cities in California

	WORK_CITY	Number of employees	Overpay	Underpay	Overpay rate
1	NEWARK	223	173	50	0.7757848
2	LOS ALTOS	117	84	33	0.7179487
3	BERKELEY	130	93	37	0.7153846
4	IRVINE	768	537	231	0.6992188
5	SANTA ANA	114	78	36	0.6842105
6	MENLO PARK	1188	812	376	0.6835017

	WORK_CITY	Number of employees	Overpay	Underpay	Underpay rate
1	San Jose	829	88	741	0.893848
2	FOLSOM	290	49	241	0.831035
3	Folsom	109	35	74	0.678899
4	San Diego	348	125	223	0.640805
5	FOSTER CITY	570	224	346	0.607018
6	San Francisco	269	111	158	0.587361

Table 10: Compare the time duration for the case status

	certified (N=140031)	certified-expired (N=3226)	certified-withdrawn (N=14146)	denied (N=4273)	withdrawn (N=5602)	Total (N=167278)
day						
Mean (SD)	17.3 (61.5)	183 (111)	356 (326)	116 (313)	135 (289)	55.6 (165)
Median [Min, Max]	6.00 [4.00, 2280]	150 [122, 1180]	254 [4.00, 1750]	6.00 [0, 2490]	2.00 [0, 1250]	6.00 [0, 2490]