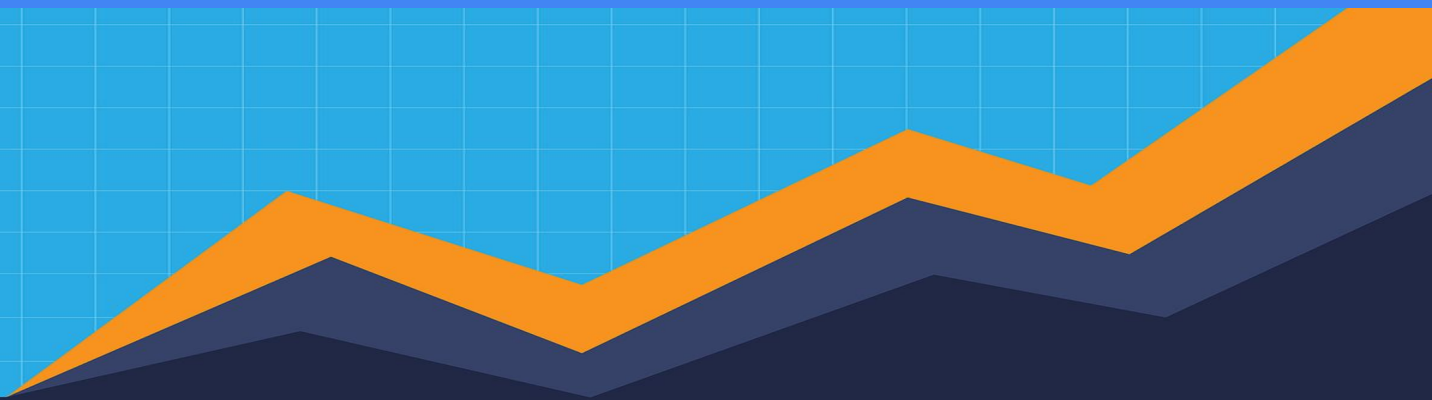# The Salary Dataset:
Data-related jobs, locations and companies

Python Charmers: Quoc Hung Nguyen, Lily Slater, Louis Perruchon, Victor Loyola, Xuefei Li

Presentation Structure

- Introduction
- I - Data Related Jobs compared to other job subgroup
- II- States with the highest salaries
- III- Comparing offered salaries to the prevailing wage
- Additional questions of the dataset

Introduction

Job Title Subgroup and Average Paid Wages Per Year

- Dataset description
  - Was conducted by the US Department of Labor's Office of Foreign Labor Certification.
  - Corporations apply work visa for foreign candidates to work in the US.
  - To make sure that US corporations will pay fair salaries to foreign candidates.
  - 8 Job Title Subgroup
  - Salaries about data-related jobs, locations and companies.
- 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.
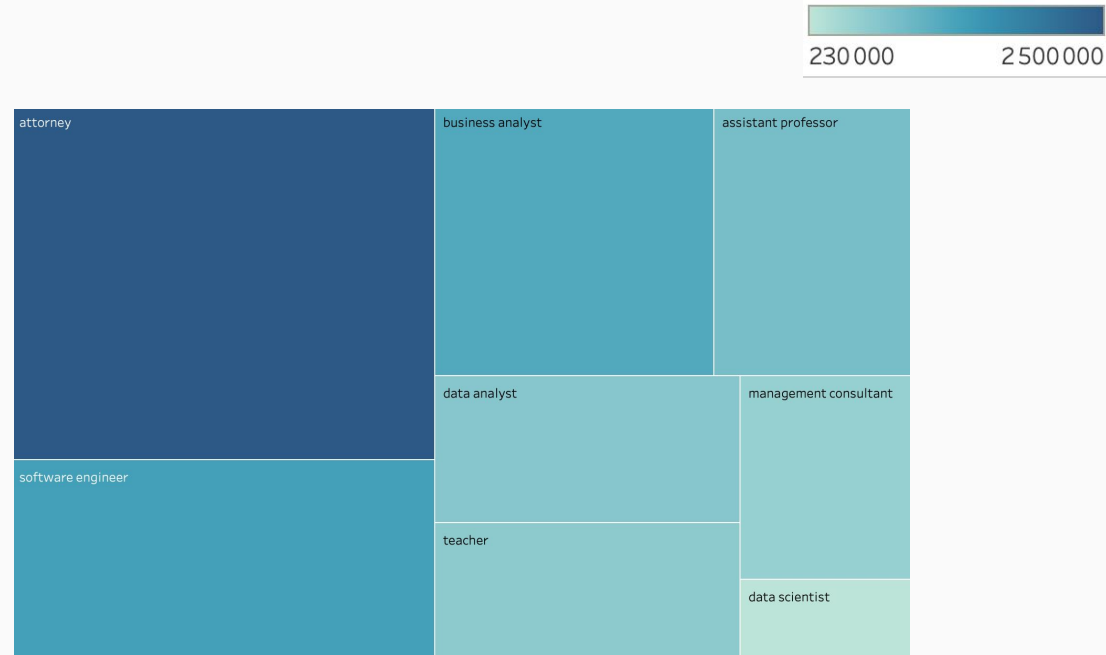
I - Data related Jobs compared to other Job subgroup

Do ecific Sub-types of data related jobs have higher or lower salaries than others?
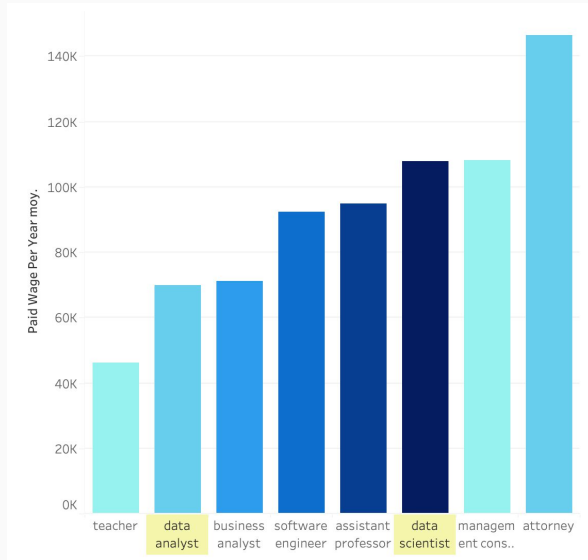
Average Wage per year grouped by Sub-types of data related jobs

Maximum Wage per year grouped by Sub-types of data related jobs

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

attorney
WINSTON & STRAWN LLP

business analyst
THE UNIVERSITY OF TEXAS
SYSTEM ADMINISTRATION

assistant professor
DIXIE STATE
COLLEGE OF UTAH

software engineer
KPI PARTNERS, INC.

data analyst
INTUIT

management
consultant
CO-CREATION
PARTNERS, INC.

teacher
FROSTBURG STATE UNIVERSITY

data scientist
NETFLIX, INC.

230 000   2 500 000
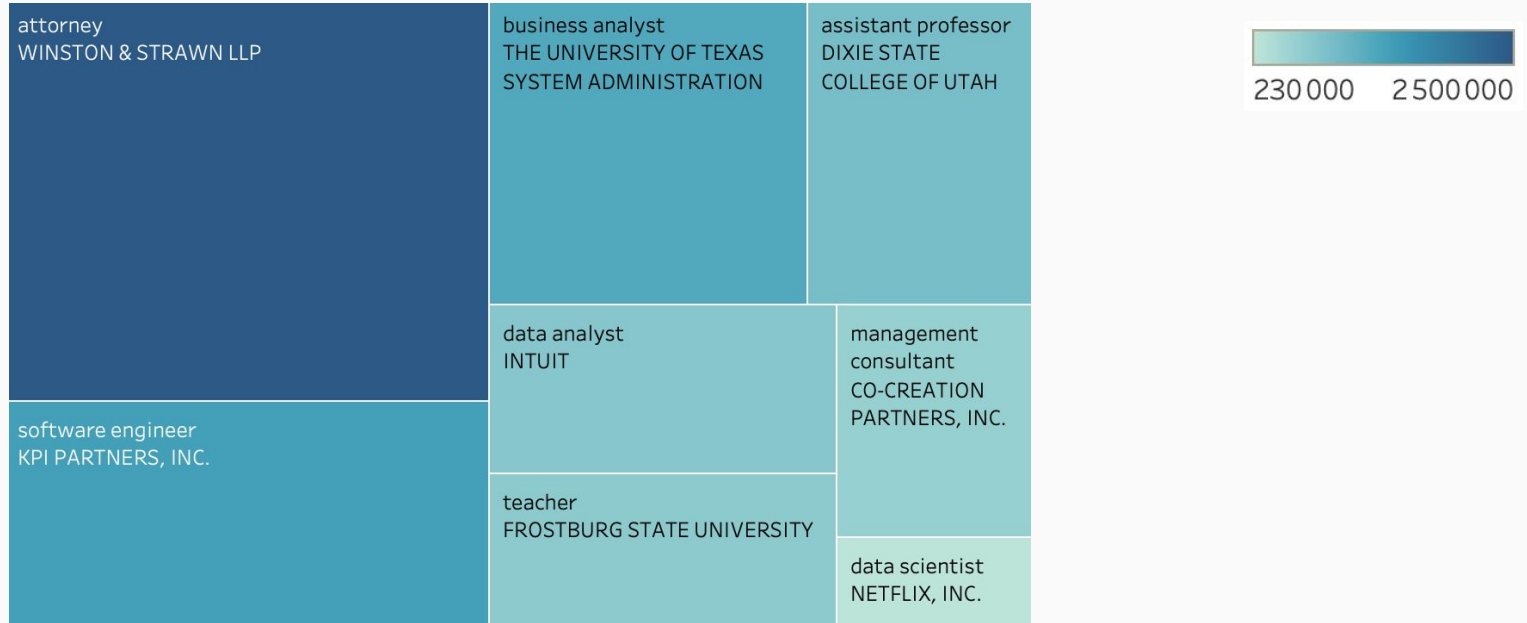
Top wage company by Subgroup Job Title

Changes with location of the job

	51013			288276

assistant professor	Kansas	109 323	management consultant	New York	116 660
	Mississippi	117 263		Connecticut	119 273
	Arizona	117 570		Georgia	120 096
	West Virginia	123 916		Pennsylvania	122 132
	Arkansas	134 445		Oklahoma	141 848
attorney	Alabama	117 451	software engineer	Maryland	85 687
	California	144 760		Massachusetts	91 577
	Oregon	156 000		New York	99 156
	New York	166 785		Washington	107 164
	District of Columbia	288 276		California	108 559
business analyst	New Mexico	73 511	teacher	Alaska	51 013
	Maryland	73 837		Connecticut	51 198
	California	76 568		California	52 278
	New York	77 574		Rhode Island	57 416
	Washington	78 092		Maryland	66 082

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

Data related jobs

Increasing top 5 average wage for each job subgroup grouped by states

Standard of living into account
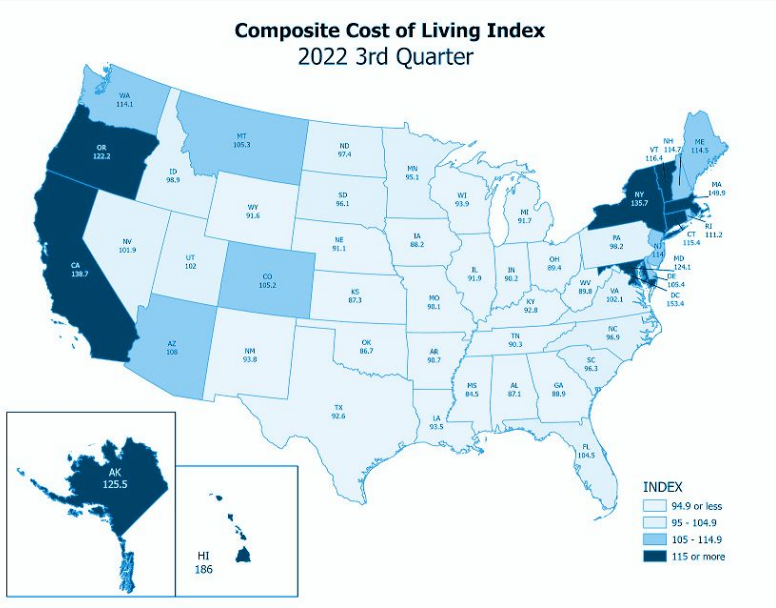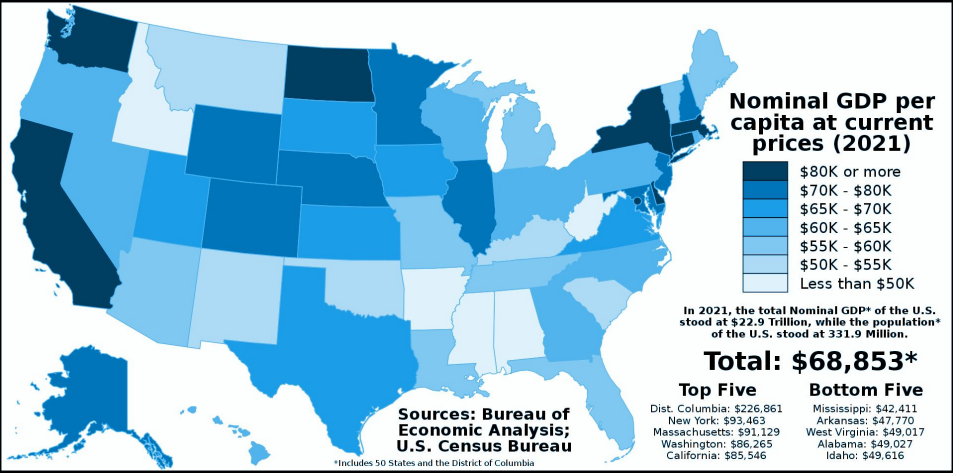
II - States with the highest Salaries for data-related jobs

Which States have the Highest Average Data-related Salaries?

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

- Louisiana, Mississippi and New Mexico were excluded for their low quality of life ratings

Differences between job sub-categories?

And how do the data-related jobs differ?

- data analyst
- data scientist

(Chart axes: Y-axis "Avg Paid Wage Per Year" with values 0, 30000, 60000, 90000, 120000; X-axis "States with Highest Avg Pay" with New York, Minnesota, Connecticut, Washington, California)
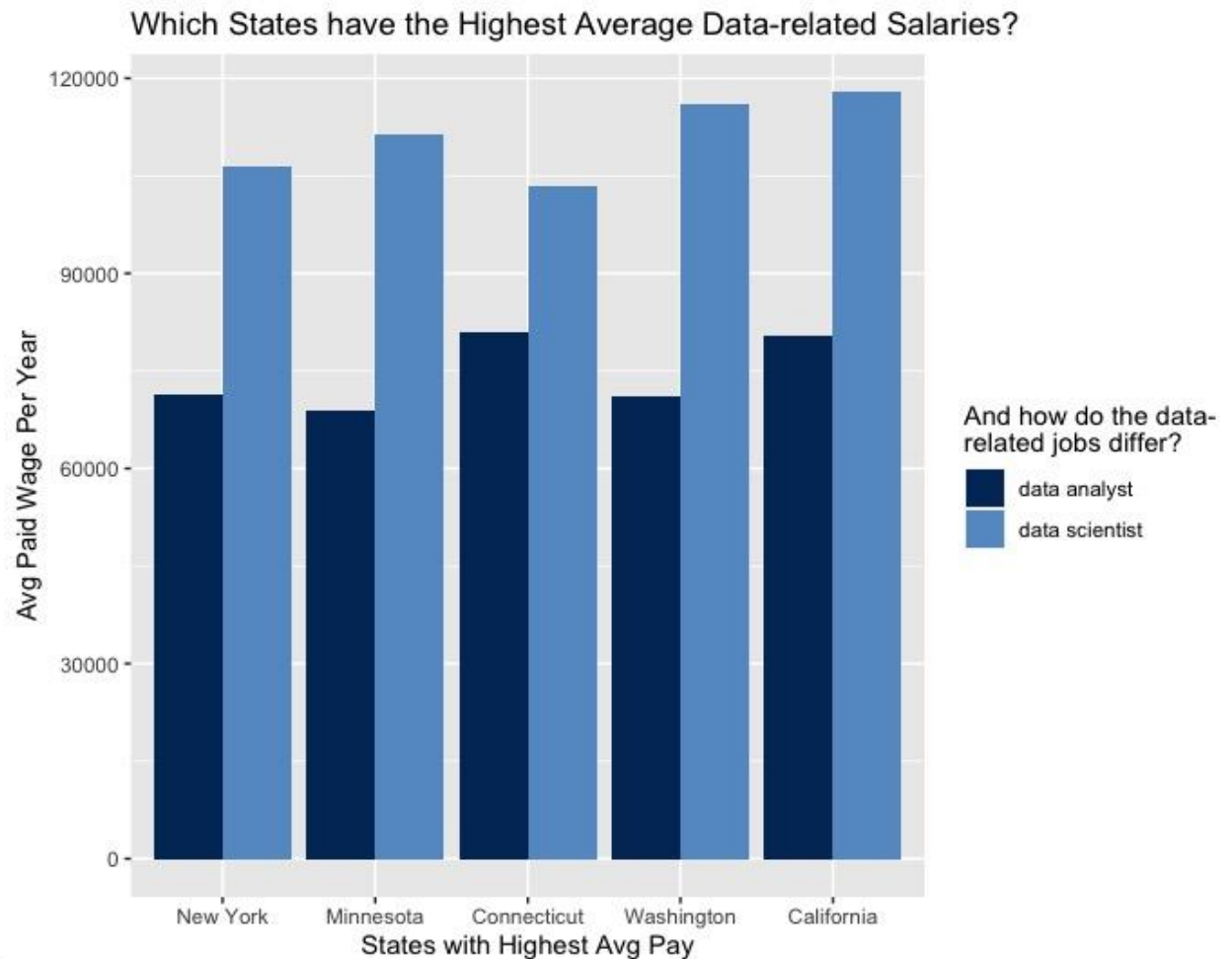
Which companies have the highest salaries for those sub-types?

Order	EMPLOYER NAME	WORK STATE	JOB TITLE SUBGROUP	PAID WAGE/YEAR
1	NETFLIX, INC.	California	data analyst	$ 220,000.00
2	AMICUS SYSTEMS INC	New York	data analyst	$ 166,400.00
3	WORLDQUANT, LLC	Connecticut	data analyst	$ 140,000.00
4	RANDSTAD TECHNOLOGIES LP	Minnesota	data analyst	$ 118,000.00
5	JPMORGAN CHASE & CO.	Washington	data analyst	$ 110,000.00
1	NETFLIX, INC.	California	data scientist	$ 230,000.00
2	EBH ENTERPRISES, INC.	New York	data scientist	$ 185,000.00
3	MICROSOFT CORPORATION	Washington	data scientist	$ 152,000.00
4	TARGET CORPORATION	Minnesota	data scientist	$ 131,789.00
5	GE CAPITAL CORPORATION	Connecticut	data scientist	$ 125,000.00

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

III - Comparing offered salaries to the prevailing Wage

Job Categories

Mean Received vs Median Received
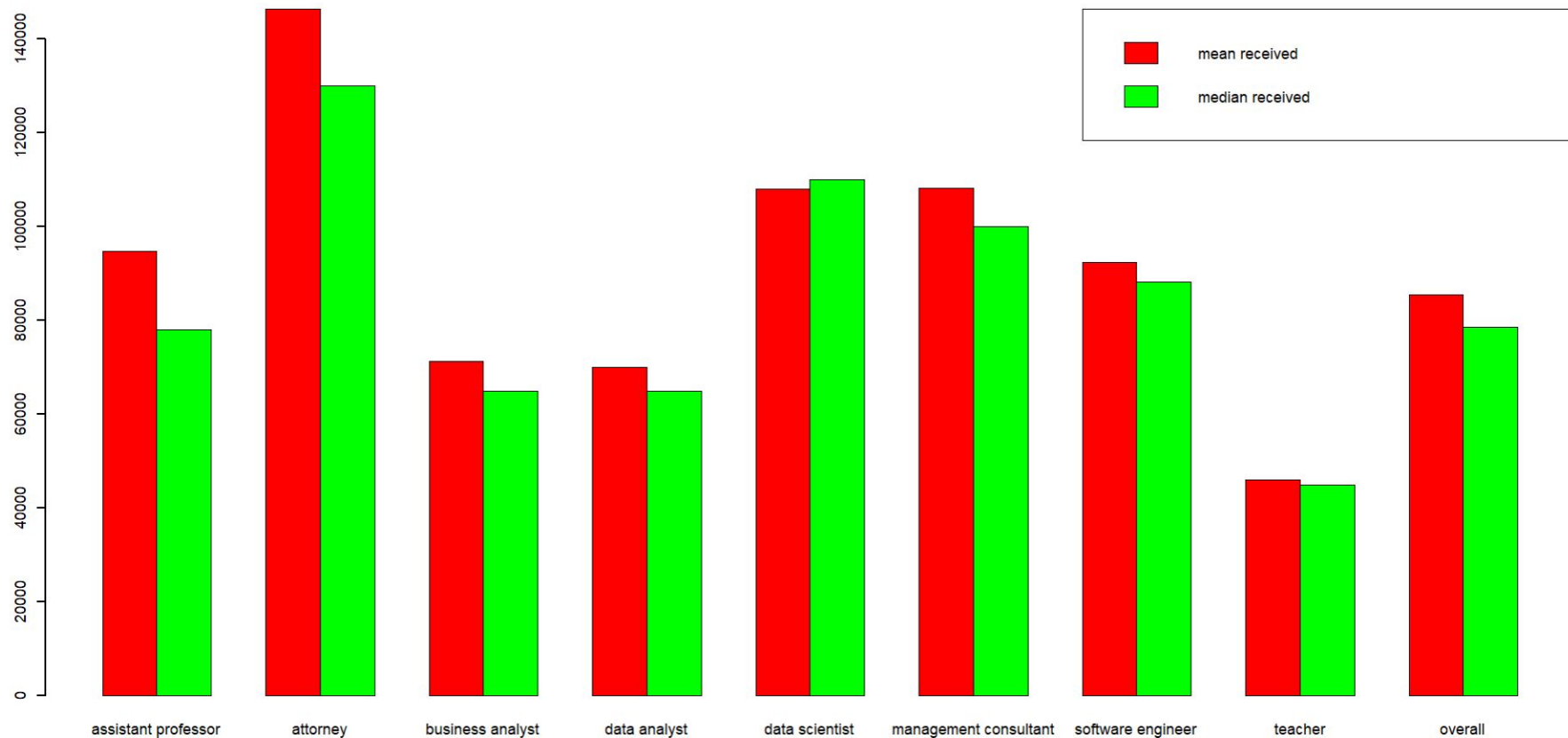
Mean Received vs mean prevailing
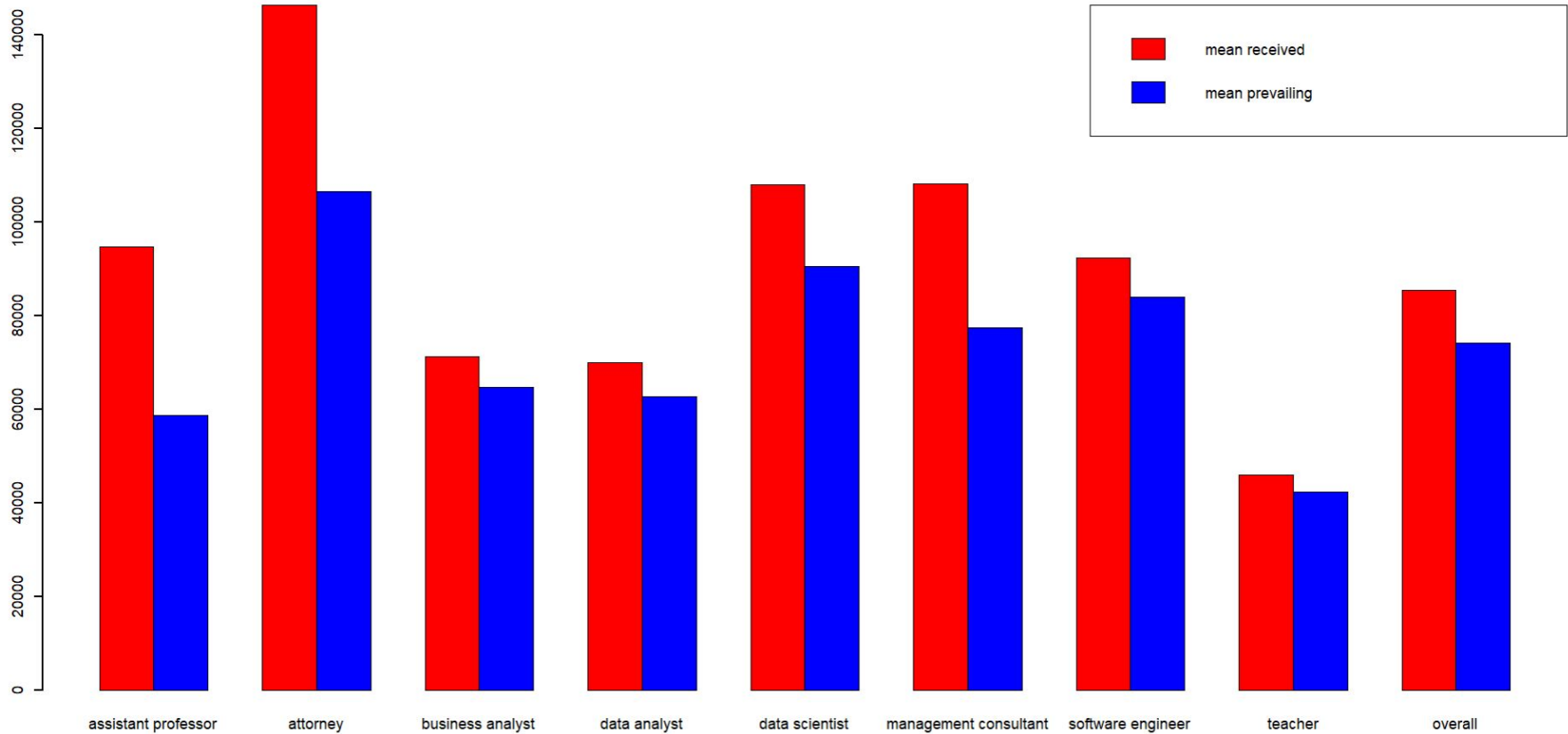
Underpay / Overpay rate

Underpay / Overpay for Job Sub - Categories

Job Sub-Category	Underpay	Overpay
Teacher	0.57	0.43
Software engineer	0.59	0.41
Management Consultant	0.51	0.49
Data Scientist	0.58	0.42
Data Analyst	0.77	0.23
Business Analyst	0.7	0.3
Attorney	0.65	0.35
Assistant Professor	0.66	0.34

■ Underpay ■ Overpay

Companies

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

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

	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

=> The cost of living does affect the overpay / underpay rate.

Additional questions of the dataset

In California, which **cities** have the highest salaries?

Californian Cities with the Highest Paid Data-related Jobs on average

Work State / Job Title Subgroup / Work City

California

data analyst | data scientist

Avg. Paid Wage Per Year

| Job Title Subgroup |
| data analyst |
| data scientist |

REDWOOD SHORES, MENLO PARK, LOS GATOS, SAN JOSE, BRISBANE, SAN RAMON, SAN BRUNO, PLEASANTON, CARLSBAD, CUPERTINO, MENLO PARK, LOS GATOS

160K, 140K, 120K, 100K, 80K, 60K, 40K, 20K, 0K
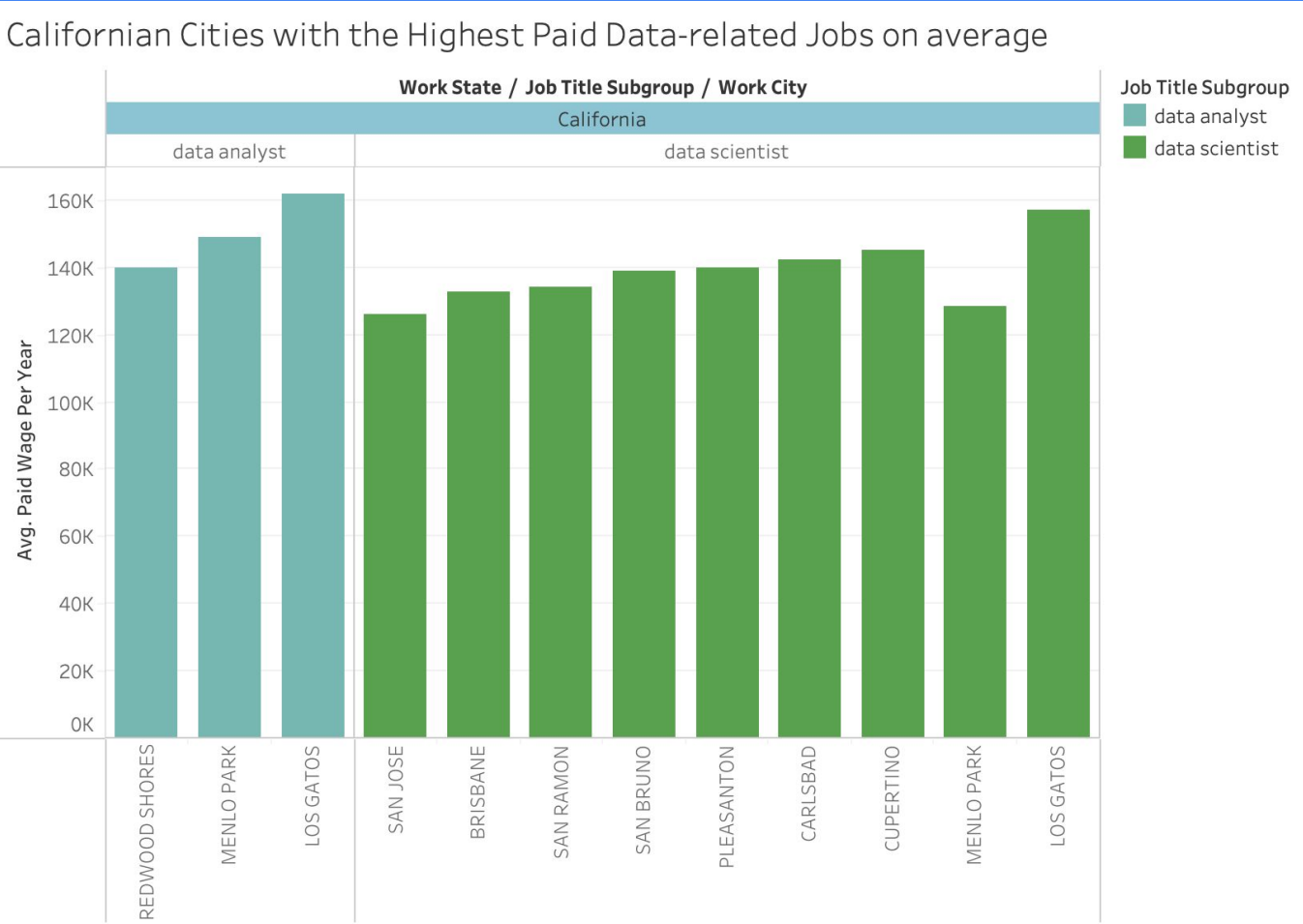
In California, which **cities** have the highest overpay/underpay rate

	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

Additional questions of the dataset

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

	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]