DataPath

Data Science Related Jobs Salary Analysis

Moch Nabil Farras Dhiya

Bandung Institute of Technology

April 20th, 2022

Agenda

Context

Analyses

Summary and Recommendations

Context

Background

The digitalization of many fields in the industry is making a lots of job went to extinct. In contrast to that, many IT-related jobs, including Data Science, are highly increasing in demand. However, companies tend to have their own preferable software in executing their daily tasks. Further analyses are then required to help DS aspirants, especially in the US, get a better chance at getting recruited in their desirable companies.

Main Objectives

1. To find the distribution of skills needed in US states' companies (limited to 38 states).

2. To find the relationship between the company state and the following aspects:

    *   Skills (Python, Excel, Hadoop, Sspark, Tableau, Big Data, Deep Learning,)

    *   Salary

    *   Workload
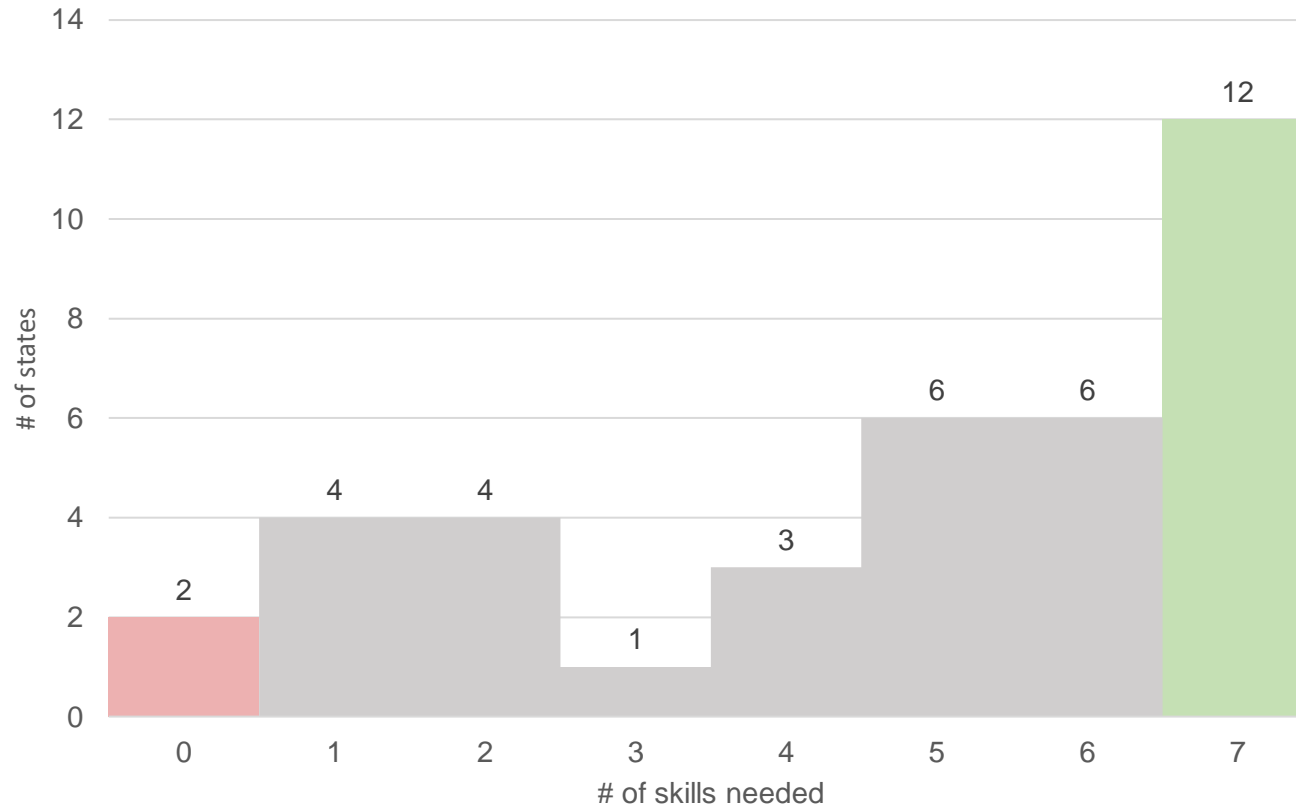
Agenda

Context

Analyses

Summary and Recommendations

The # of skills needed for DS aspirants varies (0 to 7 skills) among each state

DataPath

of skills needed distribution over state

Key Findings

- Companies in **CA**, **CO**, **DC**, **GA**, **IL**, **MA**, **MD**, **MO**, **NY**, **TN**, **TX**, and **VA** (states) require their applicants to master **all** 7 skills.

- Companies in **Mississippi** and **Rhode Island** (states) **do not** have **any** skills requirement for their DS applicants.

- On **average**, DS aspirants must master **4-5 skills** to get a better chance on landing a job.

of skills needed have a correlation with state' population

Rank	State	Population (July 1, 2021)
1	CA	39,237,836
21	CO	5,812,069
50	DC	689,545
8	GA	10,799,566
6	IL	12,671,469
15	MA	6,984,723
18	MD	6,165,129
19	MO	6,168,187
4	NY	19,835,913
16	TN	6,975,218
2	TX	29,527,941
12	VA	8,642,274

Rank	State	Population (July 1, 2021)
35	MI	2,961,279
44	RH	1,097,379

The only anomaly found in the data

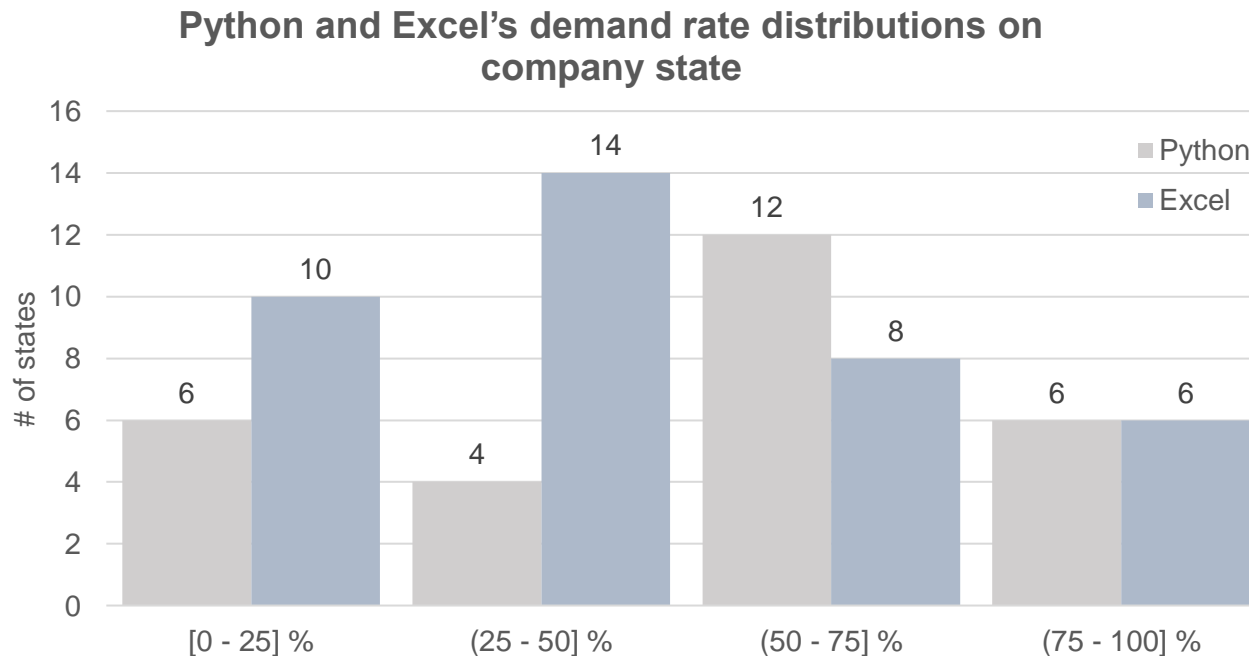**Key Findings:** States with **higher population** tend to require **more skills** as a requirement for DS jobs.

Python and Excel <u>ARE</u> high-demand skills
Overall, more than 44.76% DS jobs (of 38 states) <u>require</u> those skill(s)

Python's Demand Rate	# of States	Excel's Demand Rate	# of States
100% **(Highest)**	10	100% **(Highest)**	4
0% **(Lowest)**	5	0% **(Lowest)**	7

Python and Excel's demand rate distributions on company state

Key Findings

- Python is a **mandatory** skill for companies in **CT**, **DE**, **FL**, **GA**, **IN**, **KS**, **NH**, **OR**, **UT**, and **WV** (states).

- Excel is a **mandatory** skill for companies in **IA**, **KS**, **LA**, and **OK** (state).

- **More than 10** out of 38 states (25%) have **higher** (>50%) number of jobs that requires Python. The same also applies for Excel.

- Learning both skills **are** prioritized. Though **Python** are **more**.
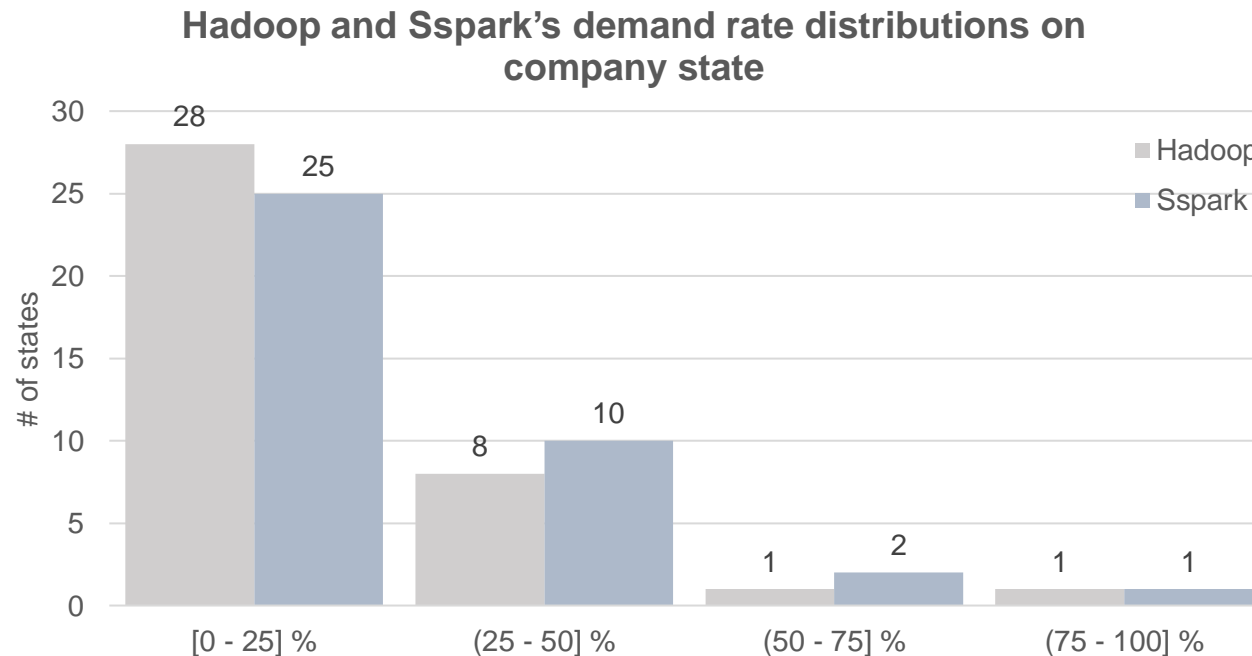
Hadoop and Sspark <u>ARE NOT</u> high-demand skills
Overall, more than 78.45% DS jobs (of 38 states) <u>do not require</u> those skill(s)

DataPath

Hadoop's Demand Rate	# of States	Sspark's Demand Rate	# of States
100% (Highest)	1	100% (Highest)	1
0% (Lowest)	14	0% (Lowest)	14

Hadoop and Sspark's demand rate distributions on company state

Key Findings

- Hadoop is a **mandatory** skill **only** for companies in **Delaware** (state).

- Sspark is a **mandatory** skill **only** for companies in **New Hampshire** (state).

- **36** out of 38 states have **lower** (≤50%) number of jobs that require Hadoop. The same also applies to Sspark.

- Learning any of those skills **are not** prioritized, **except** you want to apply to companies in Delaware or New Hampshire.
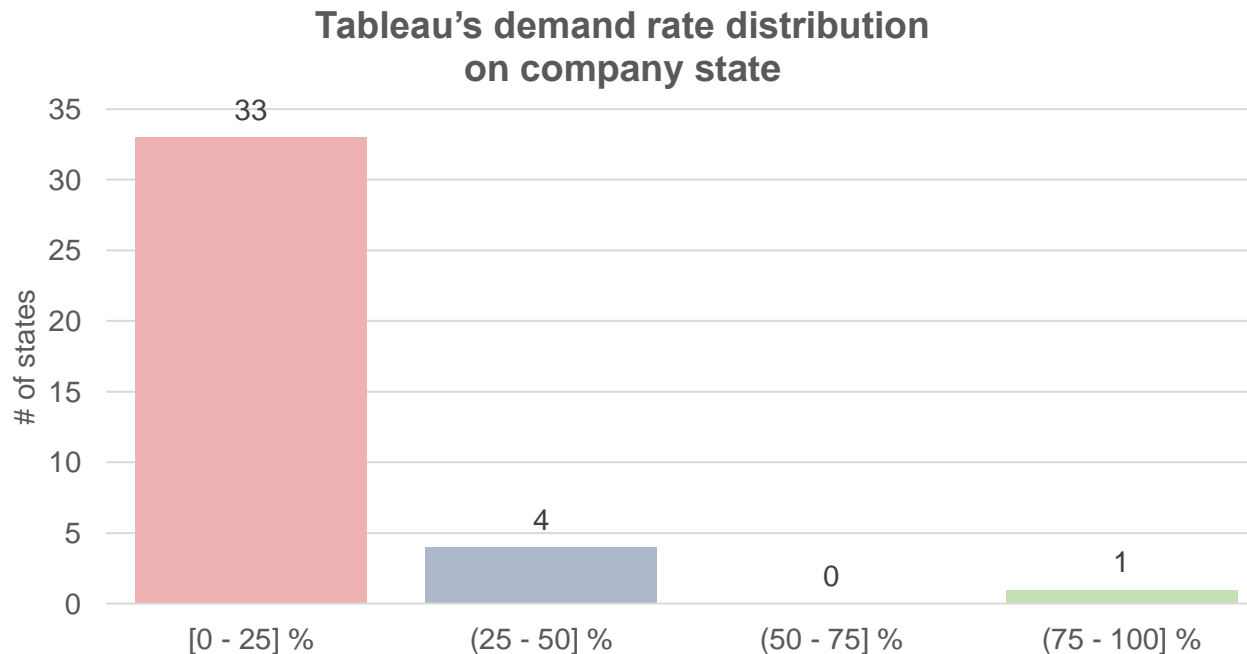
Tableau ARE NOT high-demand skills
Overall, more than 81.48% DS jobs (of 38 states) do not require Tableau skill

DataPath

Tableau's Demand Rate	# of States
100% (Highest)	1
0% (Lowest)	15

Key Findings

- Tableau is a **mandatory** skill **only** for companies in **Iowa** (state).

- **37** out of 38 states have **lower** (≤50%) number of jobs that require Tableau skill.

- **Except** you have already master other skills (especially high-demand skills), learning Tableau **is not** prioritized.
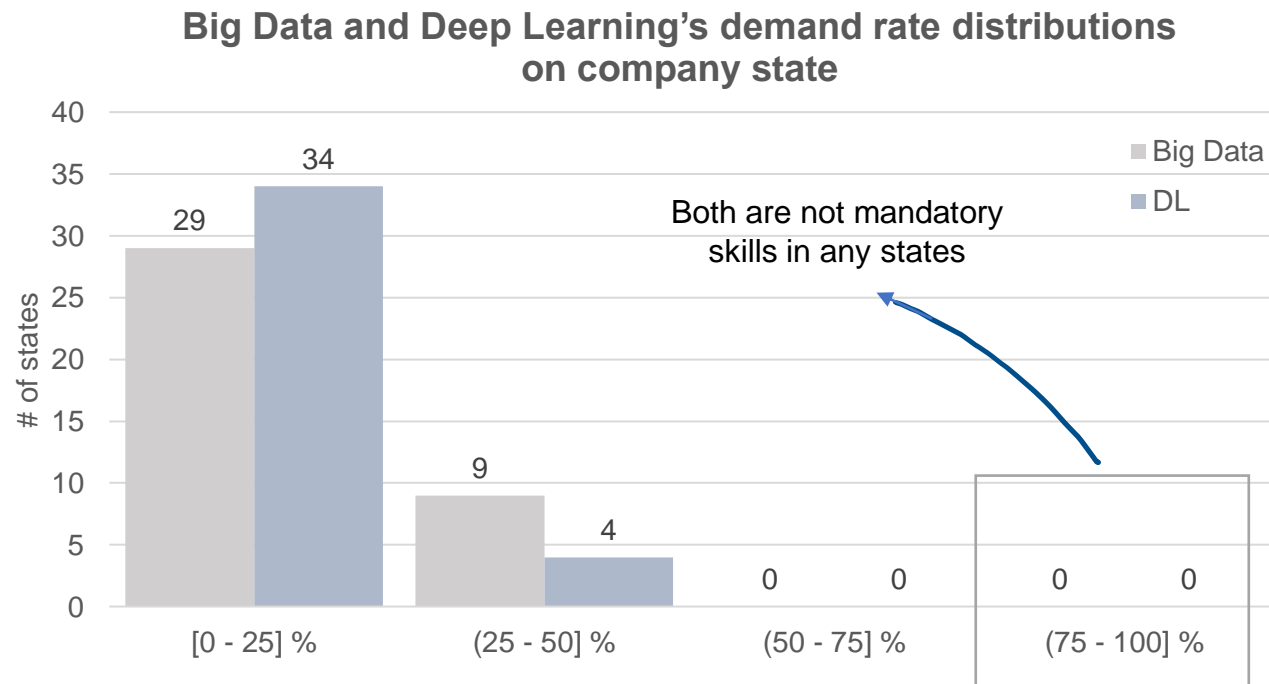
Tableau's demand rate distribution on company state

# of states

35	
30	
25	
20	
15	
10	
5	
0	

33   4   0   1

[0 - 25] %   (25 - 50] %   (50 - 75] %   (75 - 100] %

Big Data and Deep Learning 'ARE NOT' high-demand skills
Overall, more than 85.28% DS jobs (of 38 states) do not require those skill(s)

Big Data's demand rate	# of states	Deep Learning's demand rate	# of states
50% (Highest)	2	44.4% (Highest)	1
0% (Lowest)	12	0% (Lowest)	22

Big Data and Deep Learning's demand rate distributions on company state

Both are not mandatory skills in any states

Big Data
DL

# of states

29 — 34 — [0 - 25] %
9 — 4 — (25 - 50] %
0 — 0 — (50 - 75] %
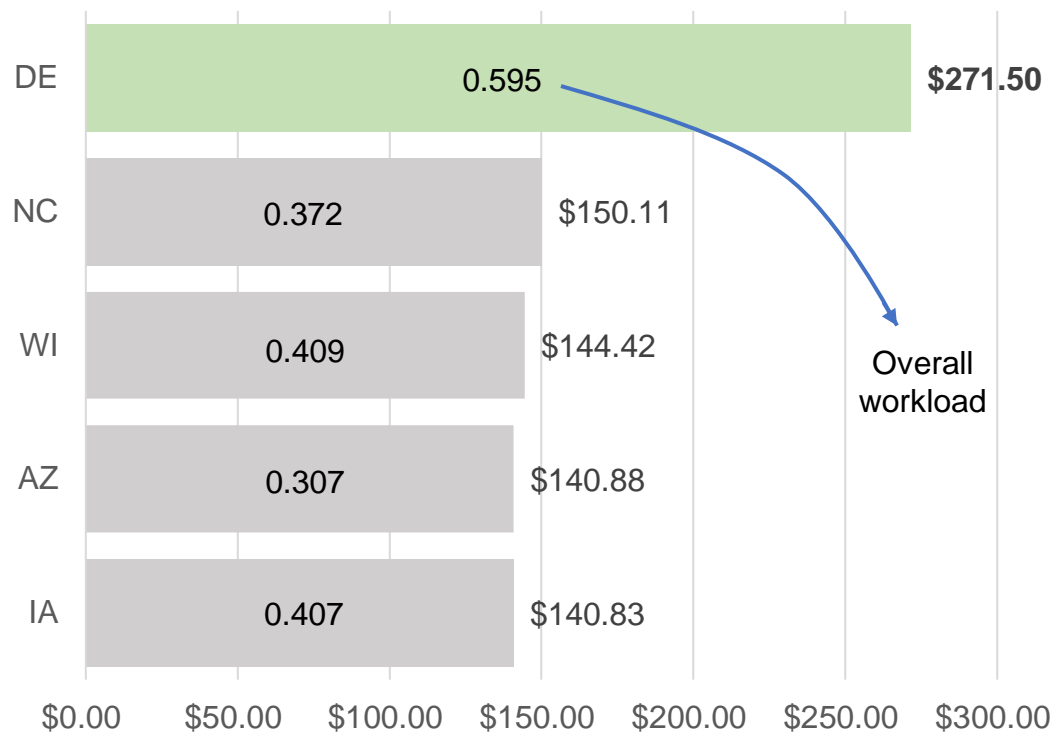0 — 0 — (75 - 100] %

Key Findings

- **Connecticut** and **New Hampshire** have the **highest** Big Data demand-rate (**50%**).

- **Georgia** has the **highest** Big Data demand-rate (**44.4%**).

- **Low-demand rate** could mean those skills are not useful for DS jobs in USA. It could also mean that it is **only** required **for senior-level**. The latter is more likely.

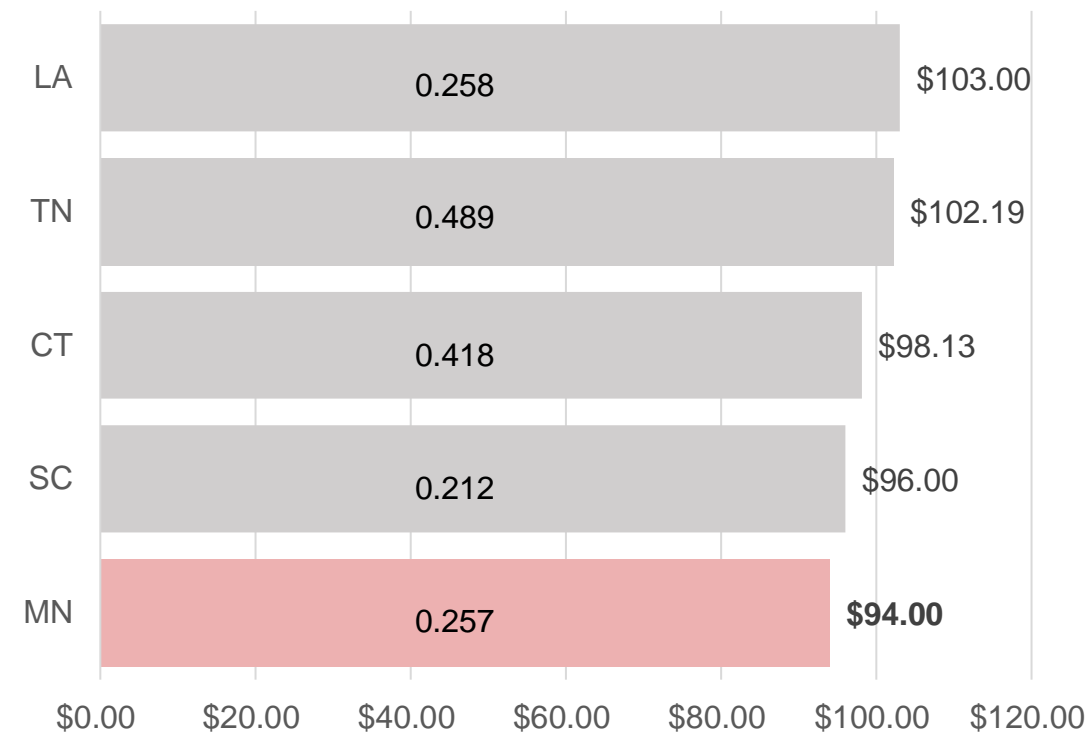- For **junior-level** DS aspirants, learning any of these skills **are not** prioritized.

Delaware has the highest average salary estimate ($271.5K), and Minnesota has the lowest average salary estimate ($94K)
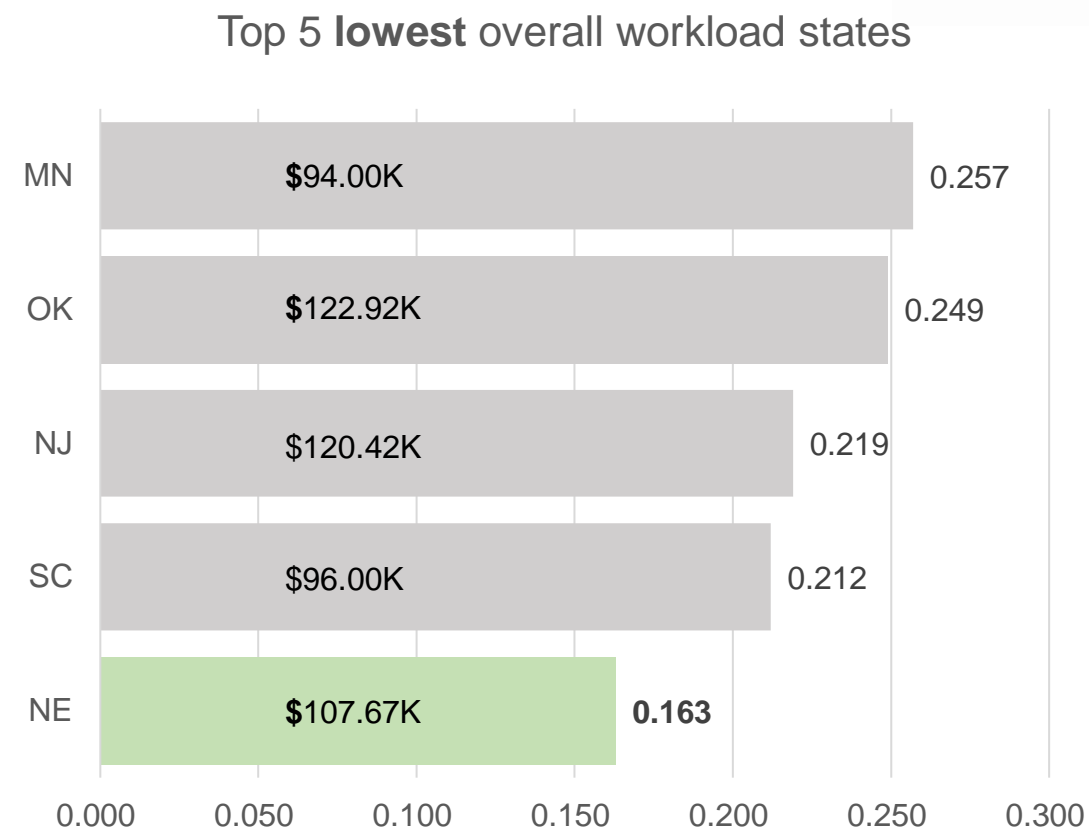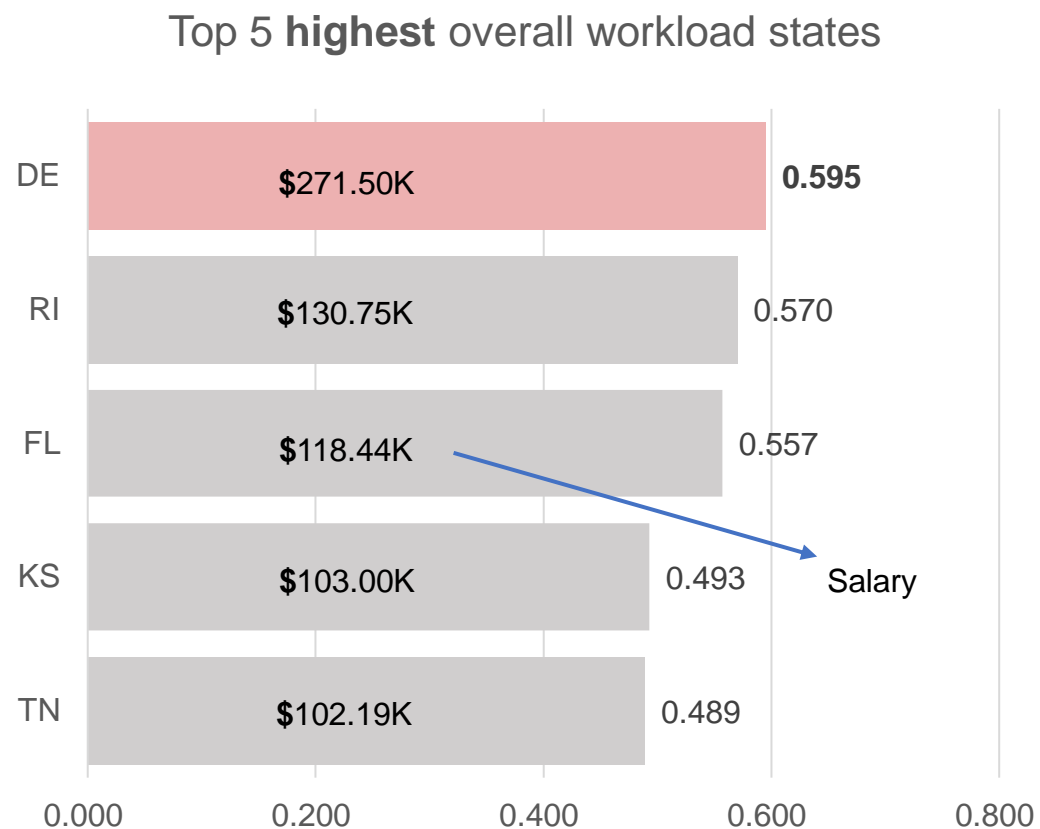
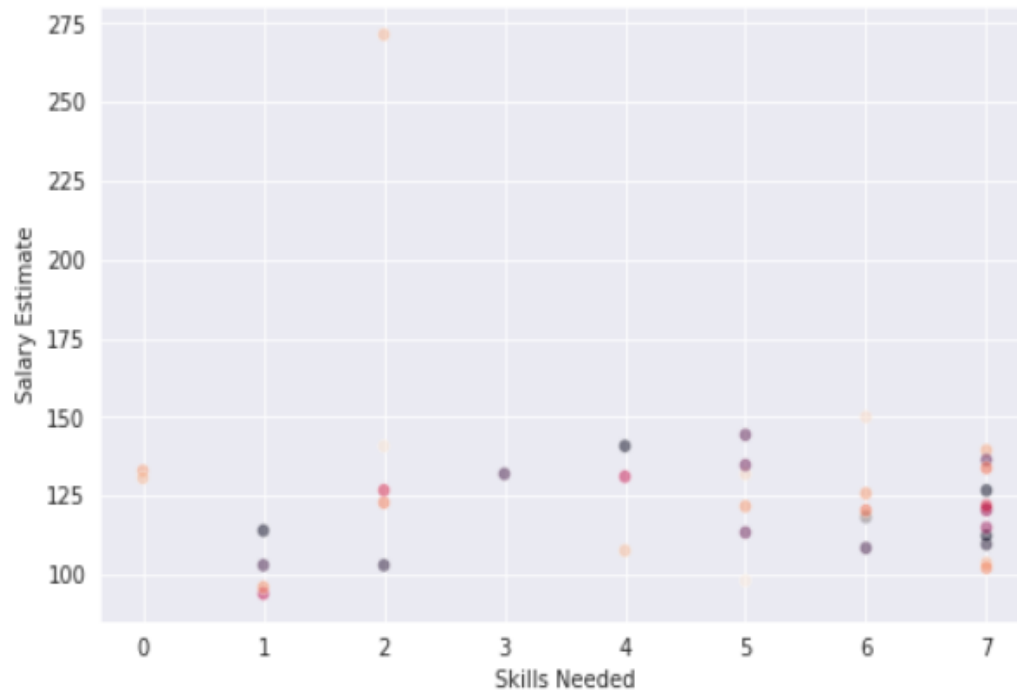## Top 5 states with the **highest** average salary

State	Workload	Salary
DE	0.595	**$271.50**
NC	0.372	$150.11
WI	0.409	$144.42
AZ	0.307	$140.88
IA	0.407	$140.83

Overall workload

$0.00 $50.00 $100.00 $150.00 $200.00 $250.00 $300.00

Top 5 states with the **lowest** average salary

State	Workload	Salary
LA	0.258	$103.00
TN	0.489	$102.19
CT	0.418	$98.13
SC	0.212	$96.00
MN	0.257	**$94.00**

$0.00 $20.00 $40.00 $60.00 $80.00 $100.00 $120.00

Delaware has the highest overall workload, and Nebraska has the lowest overall workload

Top 5 **highest** overall workload states

State	Salary	Workload
DE	$271.50K	**0.595**
RI	$130.75K	0.570
FL	$118.44K	0.557
KS	$103.00K	0.493
TN	$102.19K	0.489

Salary

Top 5 **lowest** overall workload states

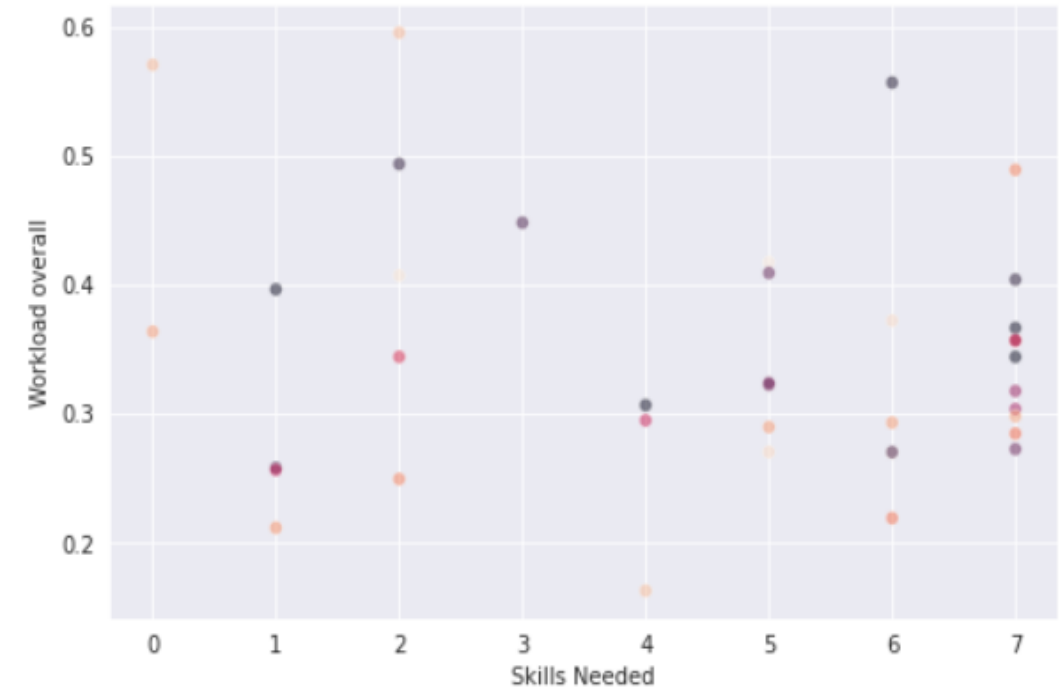| State	Salary	Workload
MN	$94.00K	0.257
OK	$122.92K	0.249
NJ	$120.42K	0.219
SC	$96.00K	0.212
NE	$107.67K	**0.163**

Hypothesis: Higher workload **does not** necessarily mean higher salary, vice versa.

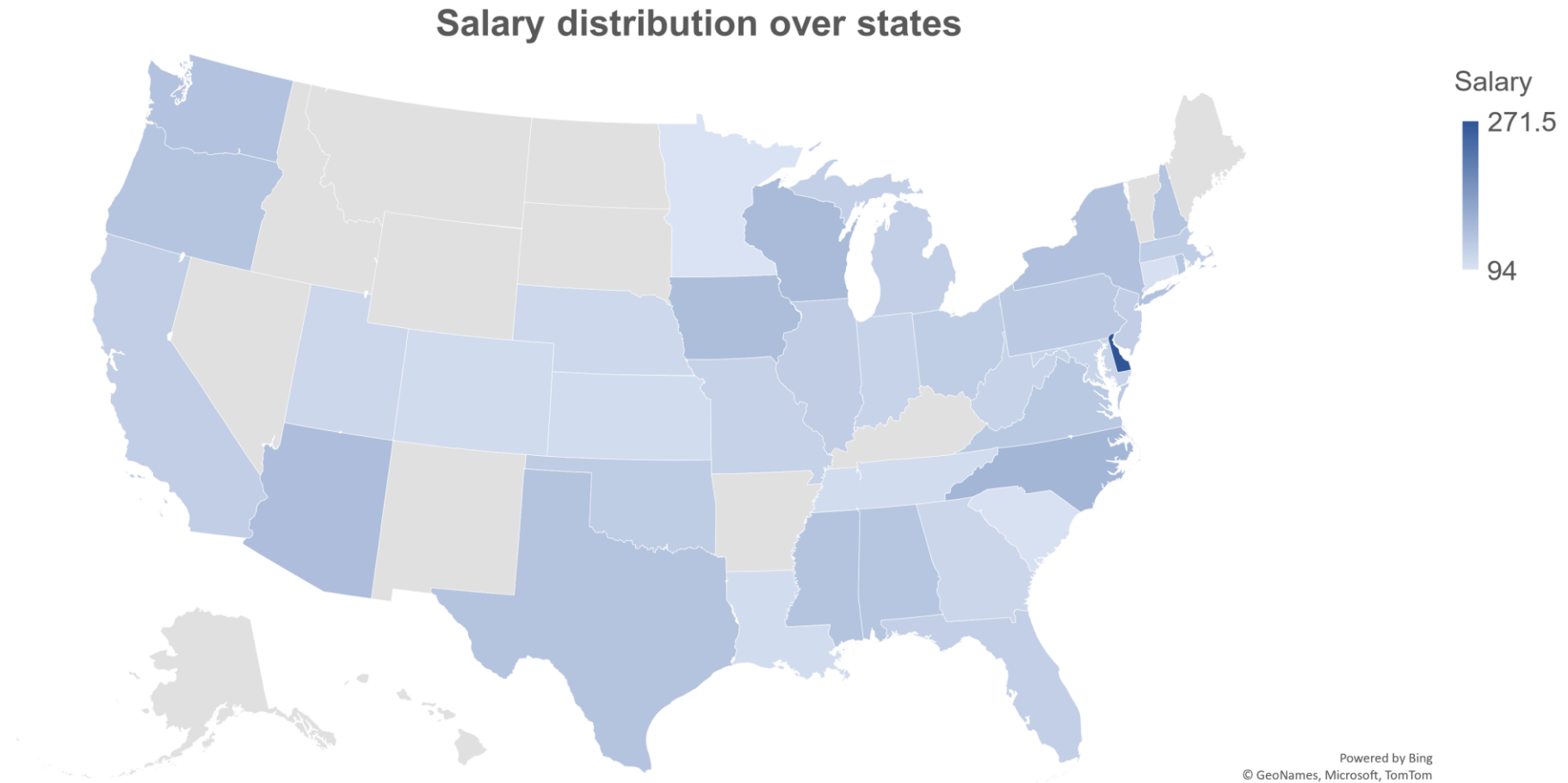# There is no correlation between each # of skills needed and overall workload with average salary estimate

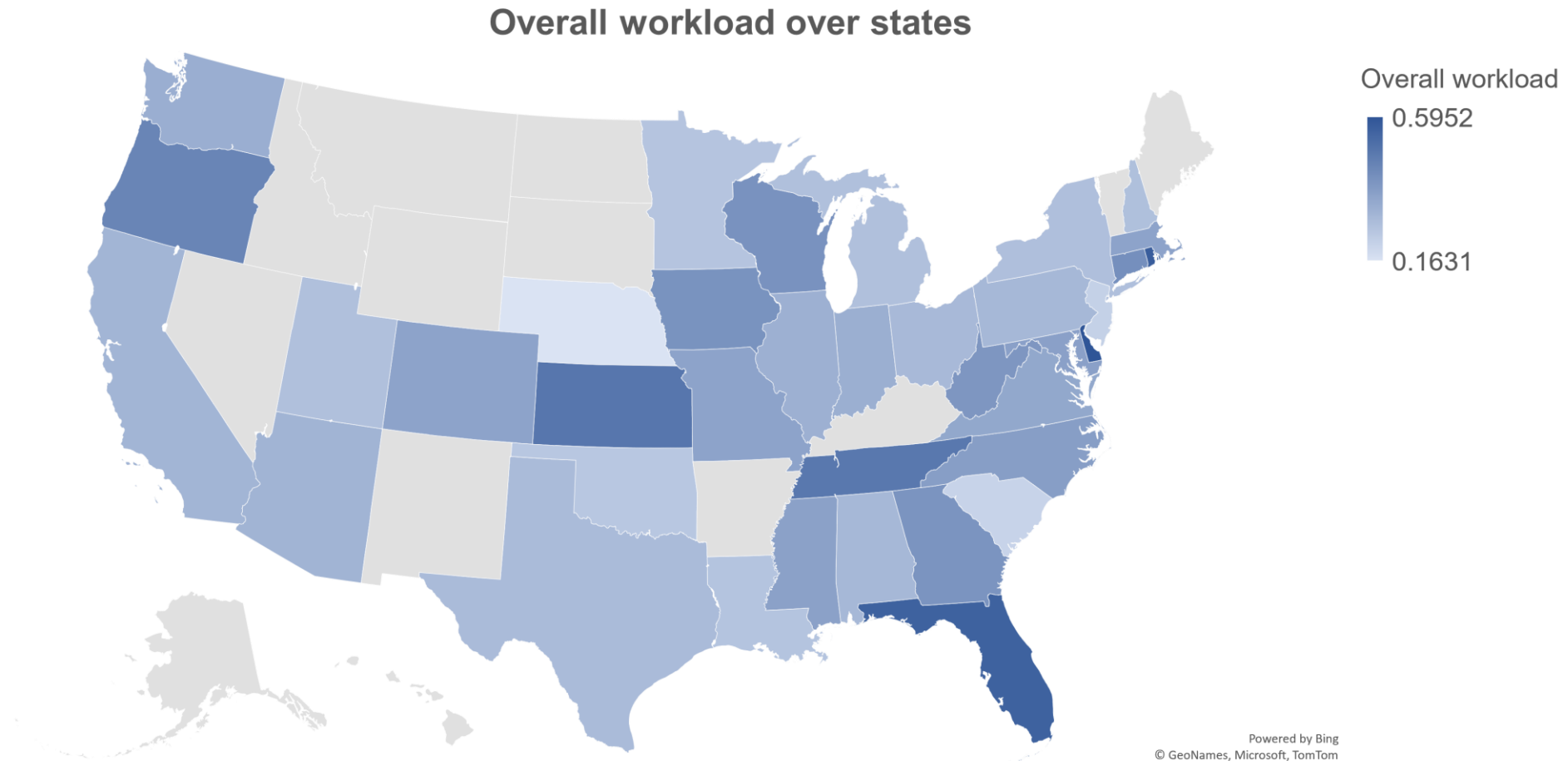**Correlation coefficient**: `-0.10215366432236693`

Correlation coefficient: `-0.14265056425899017`

Key Findings: The hypothesis is true. Thus, higher workload **does not** necessarily mean higher salary, vice versa.

There is no correlation between the company' state (geographic location) and its average salary estimate

Salary distribution over states

Salary
271.5
94

Powered by Bing
© GeoNames, Microsoft, TomTom

There is no correlation between the company' state (geographic location) and its overall workload

Overall workload over states

Overall workload

0.5952

0.1631

Powered by Bing
© GeoNames, Microsoft, TomTom

Agenda

Context

Analyses

Summary and Recommendations

Overview

Analyses' Summary	• **Skills Distribution:** The distribution of # of skills needed in companies' state varies from 0 to 7 skills. Companies in states with higher population, such as California, Texas, and New York tend to require their employees to master more skills, vice versa. • **Company State Correlations:** There are no significance correlations between company' state with salary and workload. In contrast, there is significance correlation between company' state and # of skills needed as stated above.

Recommendations

1. If you **do not** master any **high-demand skills** or aiming for **junior-level** DS jobs, then learning it right now are **highly prioritized**.
   (**Optional:** Try to apply at companies in Mississippi or Rhode Island, as it do not require any skills)
2. If you already master **Python** and **Excel** or aiming for **senior-level** DS jobs, then mastering Big Data and/or Deep Learning is **highly recommended**.
3. Applying at companies located in **Kansas**, **Tennessee**, and **Florida** are **not recommended**, since its salary are low compared to its high workload (relatively).
4. Applying at companies located in **Oklahoma** and **New** Jersey are **highly recommended**, since its salary are high compared to its low workload (relatively).