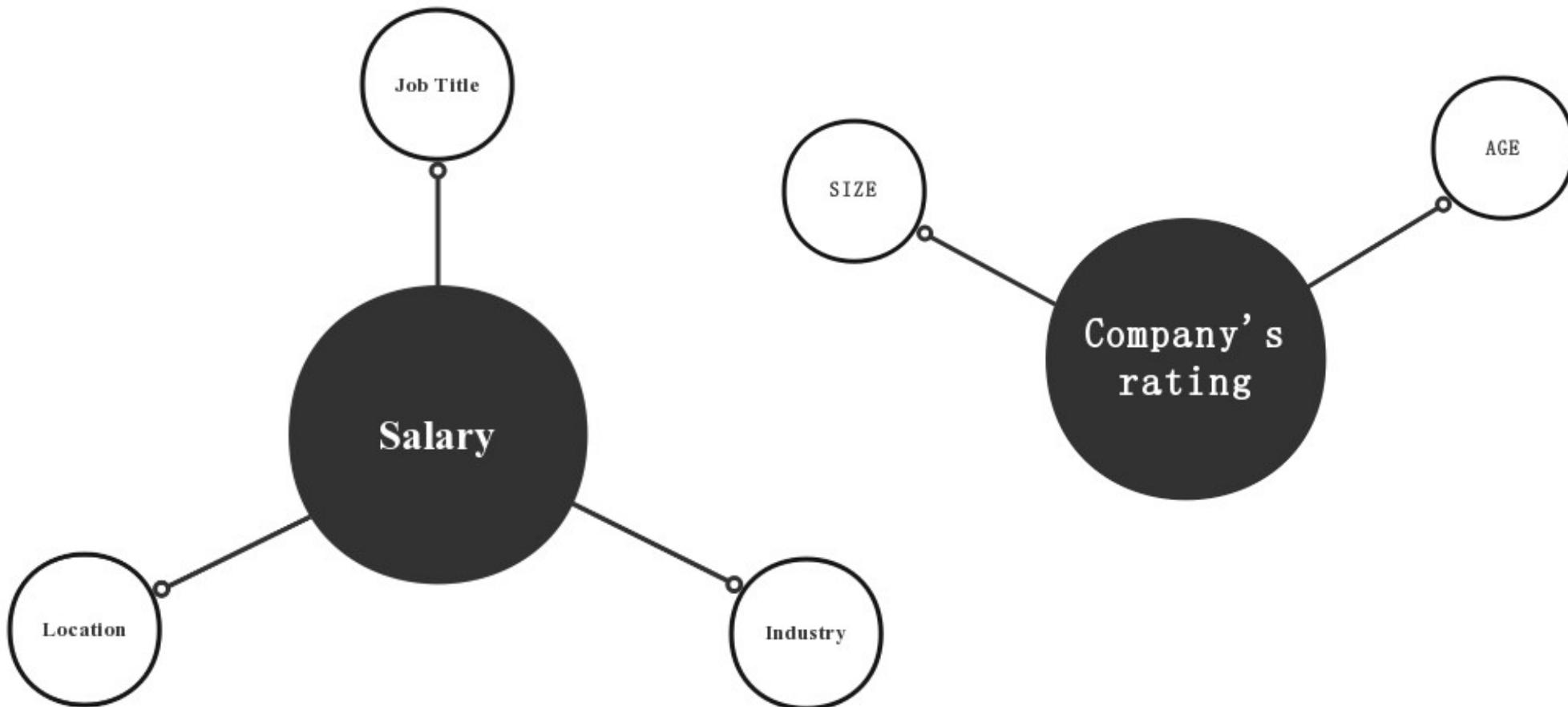


Analysis of Salary for Data Science Jobs

By Youjin An <

Problem Statement



Dataset

01

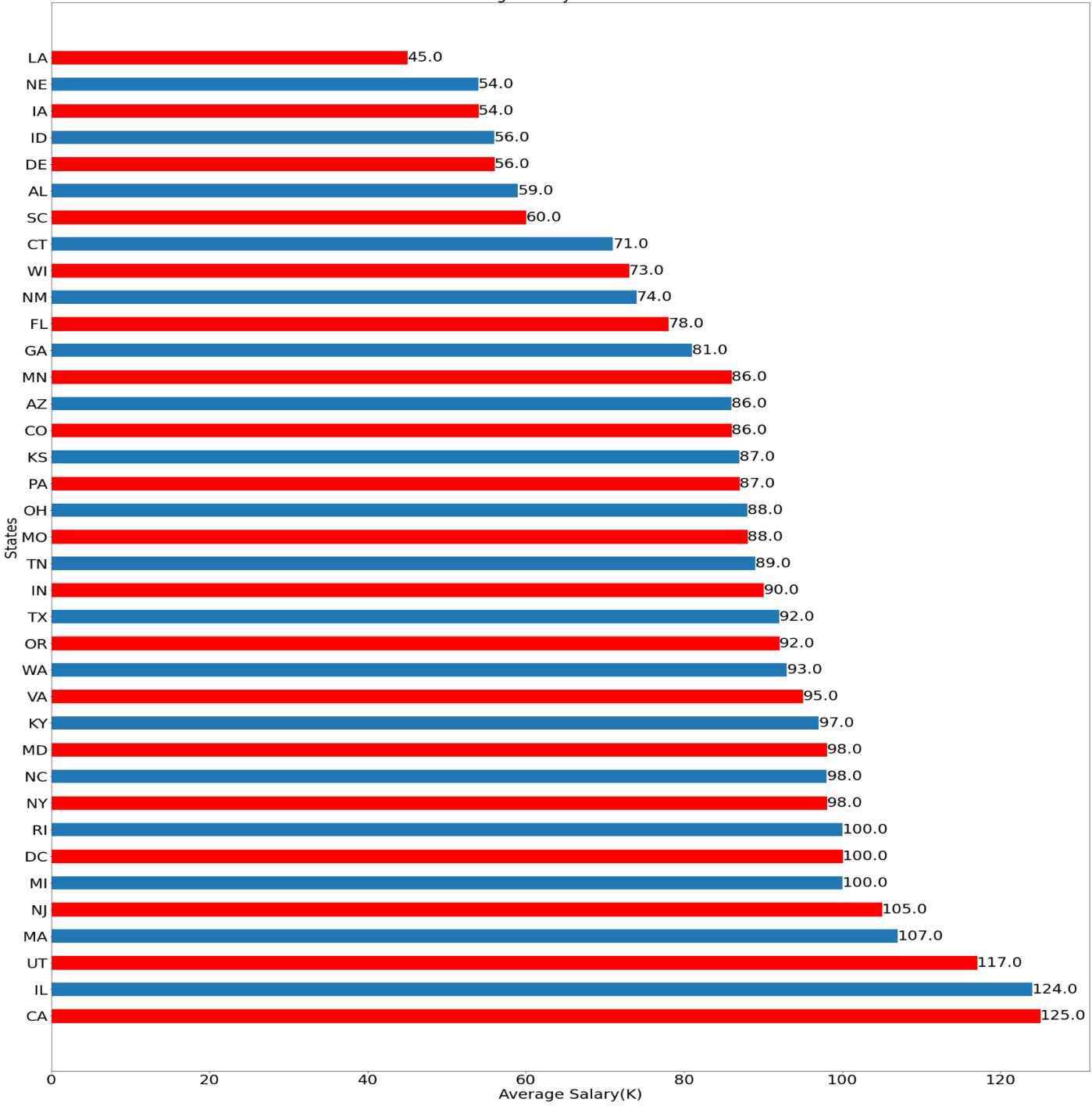
Sources
Kaggle
Glassdoor

02

Categorical
State
Industry
Job titles

Index	Rating	Founded	Ava Salary(K)	Python	spark	aws	excel	sql	sas	keras	pytorch	scikit	tensor	hadoop	tableau	bi	flink	mongo	google_an	company_age	avg_company_size
Rating	1	0.0744...	0.0503523	0.13...	0.1...	0...	-0...	0...	-...	0.0...	0.054...	0.1...	0.09...	0.148...	-0.0...	...	0...	0.00...	-0.0422...	-0.0744777	-0.147228
Founded	0.07...	1	-0.017626	0.14...	0.1...	0...	-0...	0...	-...	0.0...	0.045...	0.0...	0.06...	0.073...	0.08...	...	0...	-0.0...	0.06979...	-1	-0.567564
Avg Salary(K)	0.05...	-0.017...	1	0.315645	0.1423...	0.31...	0.1...	0...	-0...	0...	0.045...	0.1...	0.16...	0.115...	-0.0...	...	0...	0.08...	-0.1265...	0.017626	0.0988066
Python	0.13...	0.1423...	0.315645	1	0.3...	0...	0.0...	0...	0...	0.1...	0.102...	0.2...	0.20...	0.2287	0.13...	...	0...	0.09...	-0.0332...	-0.142382	-0.0991648
spark	0.15...	0.1354...	0.1719	0.33...	1	0...	-0...	0...	-...	0.1...	0.040...	0.2...	0.25...	0.509...	-0.0...	...	0...	0.22...	-0.0537...	-0.135493	-0.0888995
aws	0.08...	0.0466...	0.173971	0.18...	0.2...	1	-0...	0...	-...	0.1...	0.096...	0.0...	0.11...	0.262...	-0.0...	...	0...	0.20...	-0.0800...	-0.0466996	-0.109486
excel	-0.0...	-0.048...	-0.0805838	0.01...	-0...	-0...	1	0...	0...	-0...	-0.05...	-0...	-0.0...	0.023...	0.15...	...	-0...	0.09...	0.120017	0.048761	-0.016927
sql	0.10...	0.1446...	0.00970533	0.35...	0.2...	0...	0.1...	1	0...	0.0...	-0.00...	0.1...	0.04...	0.260...	0.37...	...	-0...	0.19...	0.05757...	-0.144625	-0.12044
sas	-0.0...	-0.011...	0.104331	0.08...	-0...	-0...	0.0...	0...	1	-0...	-0.05...	0.0...	-0.0...	-0.07...	0.15...	...	-0...	0.10...	0.09257...	0.0112068	0.12806
keras	0.08...	0.0704...	0.120863	0.18...	0.1...	0...	-0...	0...	-...	1	0.316...	0.4...	0.53...	0.012...	-0.0...	...	-0...	0.05...	0.02368...	-0.0704491	-0.0355702
pytorch	0.05...	0.0458...	0.0453792	0.10...	0.0...	0...	-0...	-...	-...	0.3...	1	0.3...	0.62...	-0.03...	-0.1...	...	-0...	-0.0...	-0.0346...	-0.0458419	0.000122787
scikit	0.12...	0.0904...	0.178549	0.24...	0.2...	0...	-0...	0...	0...	0.4...	0.374...	1	0.51...	0.146...	-0.0...	...	0...	0.03...	-0.0017...	-0.0904184	-0.0361468
tensor	0.09...	0.0676...	0.165139	0.20...	0.2...	0...	-0...	0...	-...	0.5...	0.628...	0.5...	1	-0.00...	-0.1...	...	0...	-0.0...	-0.0142...	-0.067625	-0.0268019
hadoop	0.14...	0.0738...	0.115944	0.22...	0.5...	0...	0.0...	0...	-...	0.0...	-0.03...	0.1...	-0.0...	1	0.07...	...	0...	0.18...	-0.0357...	-0.0738601	-0.00857002
tableau	-0.0...	0.0852...	-0.0780524	0.13...	-0...	-0...	0.1...	0...	0...	-0...	-0.12...	-0...	-0.1...	0.074...	1	...	-0...	0.13...	0.08161	-0.0852195	-0.0856402
bi	0.01...	0.0163...	-0.0542493	0.15...	-0...	0...	0.0...	0...	0...	-0...	-0.06...	0.0...	-0.0...	-0.00...	0.45...	1	-0...	0.00...	0.03528...	-0.0163935	-0.0444742
flink	0.05...	0.0683...	0.0877723	0.06...	0.2...	0...	-0...	-...	-...	-0...	-0.02...	0.0...	0.12...	0.144...	-0.0...	...	1	0.02...	-0.0176...	-0.0683654	0.00252283
mongo	0.00...	-0.039...	0.0851646	0.09...	0.2...	0...	0.0...	0...	0...	0.0...	-0.05...	0.0...	-0.0...	0.185...	0.13...	...	0...	1	0.01324...	0.0399543	0.0254494
google_an	-0.0...	0.0697...	-0.126539	-0.0...	-0...	-0...	0.1...	0...	0...	0.0...	-0.03...	-0...	-0.0...	-0.03...	0.08...	...	-0...	0.01...	1	-0.0697929	-0.0636757
company_age	-0.0...	-1	0.017626	-0.1...	-0...	-0...	0.0...	-	0...	-0...	-0.04...	-0...	-0.0...	-0.07...	-0.0...	...	-0...	0.03...	-0.0697...	1	0.567564
avg_company_size	-0.1...	-0.567...	0.0988066	-0.0...	-0...	-0...	-0...	-	0...	-0...	0.000...	-0...	-0.0...	-0.00...	-0.0...	...	0...	0.02...	-0.0636...	0.567564	1

Average Salary for Each State



Hypothesis 1:

01

There is a big difference in the average salary of data science jobs in different states.

Method :

02

Horizontal Bar Plot

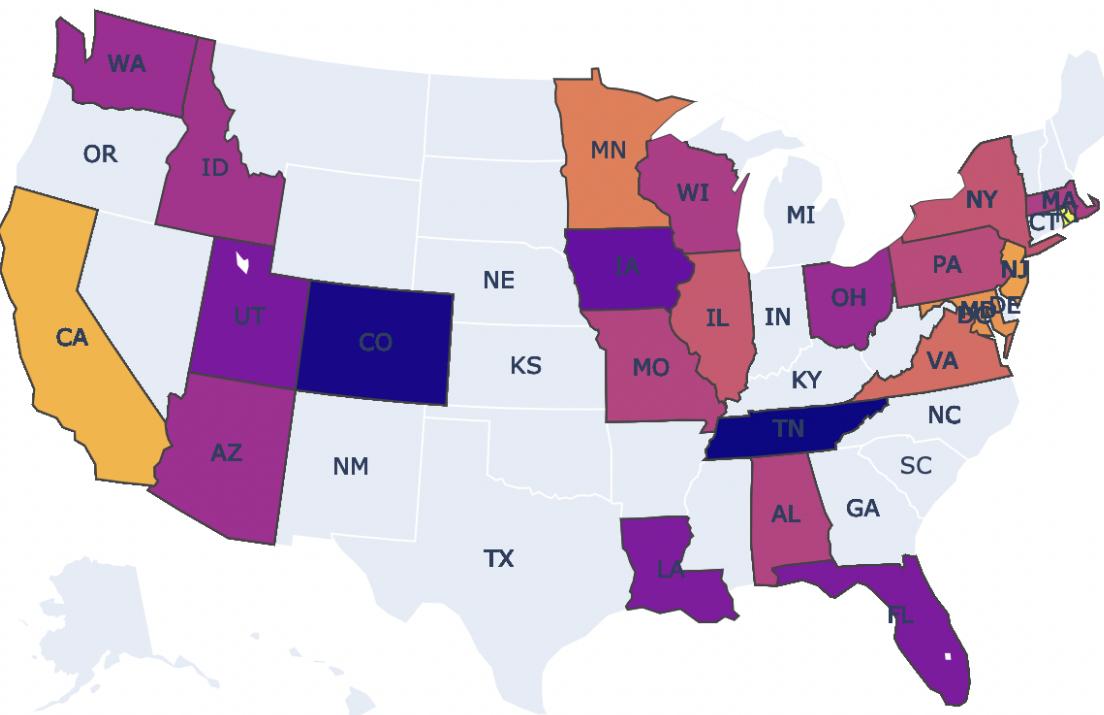
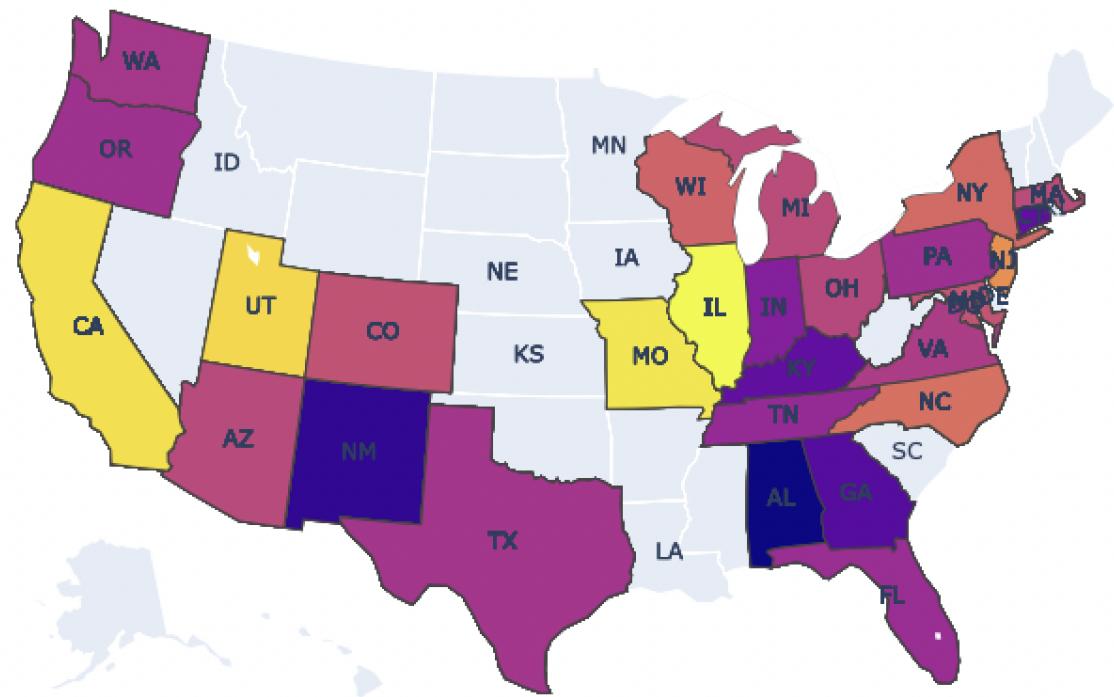
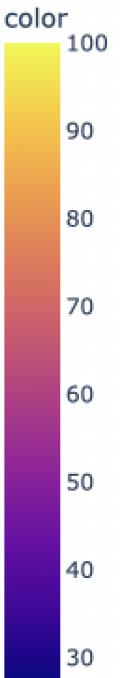
Results :

03

1. Relatively low: LA - SC
2. Medium: CT - MI
3. Relatively high: NJ - CA
4. No huge difference

Choropleth Map

data scientist
other scientist
data engineer
analyst
machine learning engineer
Data scientist project manager
data analytics
data modeler
director

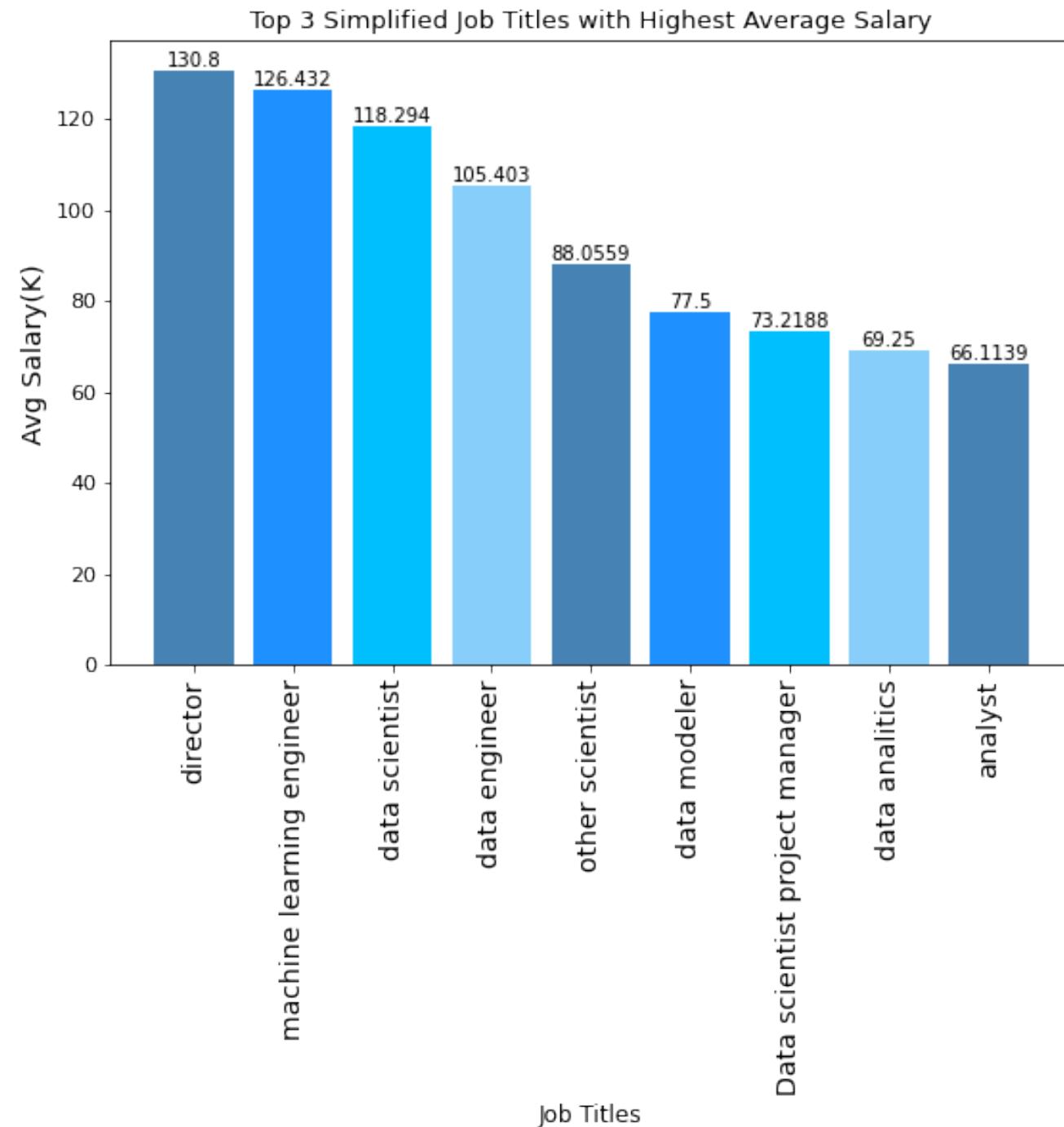


Enter job title you are looking for:
data scientist

Enter job title you are looking for:
analyst

Hypothesis 2:

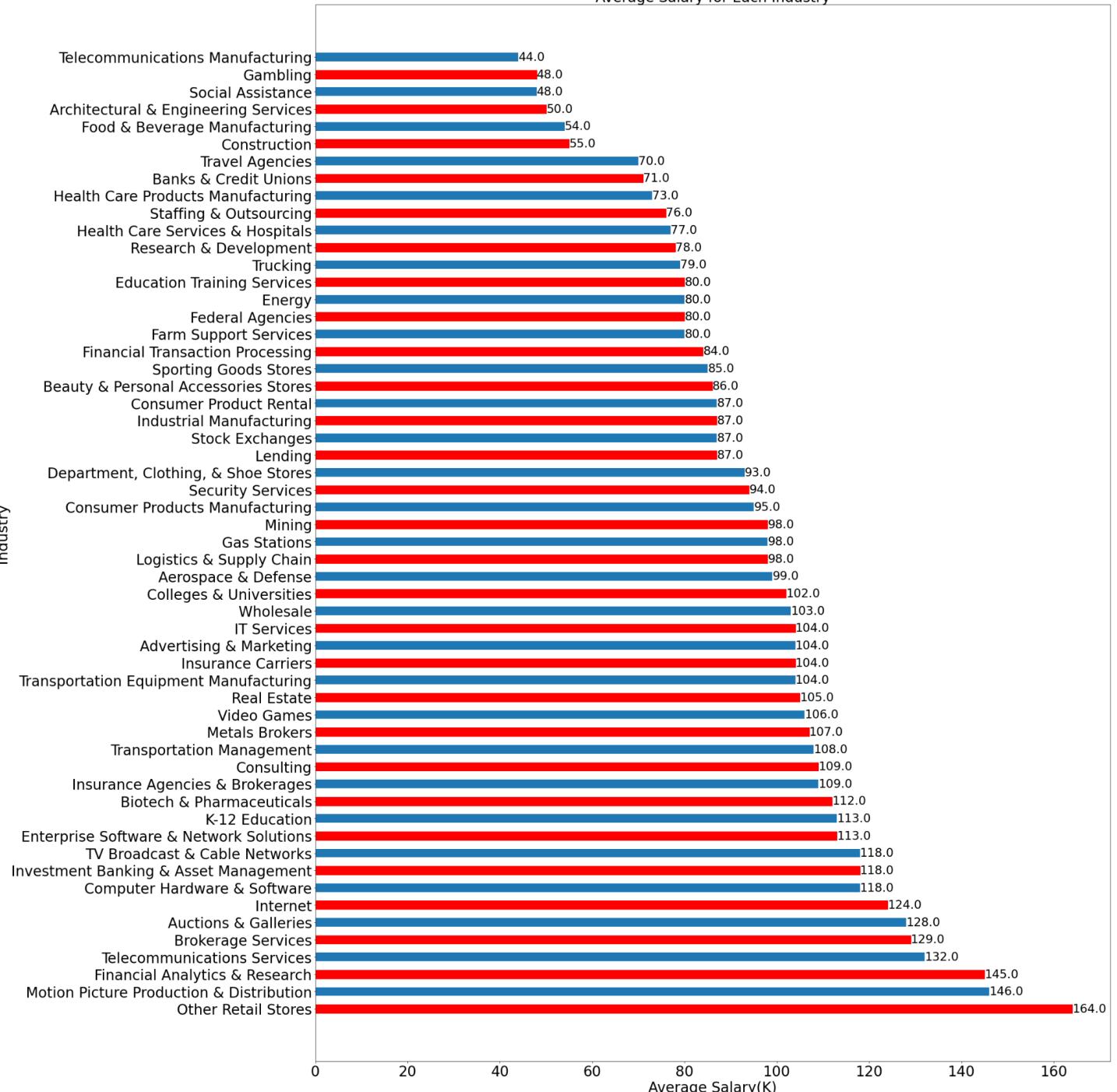
**Director, Machine Learning
Engineer, and Data Scientist
Project Manager are the top
3 jobs with the highest
average salary.**



Probability of Having Skills for Different Positions

Index	Python	spark	aws	excel	sal	sas	keras	pvtorch	scikit	tensor	hadoop	tableau	bi	flink	mongo	google an	an ▲
director	0	0	0.2	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0
machine learning engineer	0.82	0.18	0.27	0.32	0.59	0	0	0.23	0.32	0.41	0.18	0	0	0	0	0	0
data engineer	0.67	0.6	0.52	0.41	0.73	0.01	0	0	0	0.02	0.44	0.08	0.03	0.06	0.12	0	0
other scientist	0.09	0	0.11	0.44	0.02	0.02	0	0.01	0	0.01	0	0	0	0	0	0	0
data modeler	0.4	0.2	0.4	0.8	0.8	0	0	0	0	0	0.2	0	0.4	0	0	0	0
Data scientist project manager	0.2	0	0.07	0.73	0.67	0	0	0	0	0	0	0.8	0.13	0	0	0	0
data analitics	0.62	0	0	0.5	0.5	0.25	0	0	0	0	0	0.5	0.25	0	0	0	0
data scientist	0.77	0.27	0.24	0.5	0.56	0.17	0.09	0.11	0.15	0.19	0.19	0.25	0.09	0.01	0.06	0.01	0
analyst	0.32	0.06	0.08	0.76	0.74	0.11	0	0	0	0	0.03	0.4	0.18	0	0.03	0.11	0

Average Salary for Each Industry



01

Hypothesis 3:
The average salary level
depends on the industry.

02

Method :
Bar Plot

03

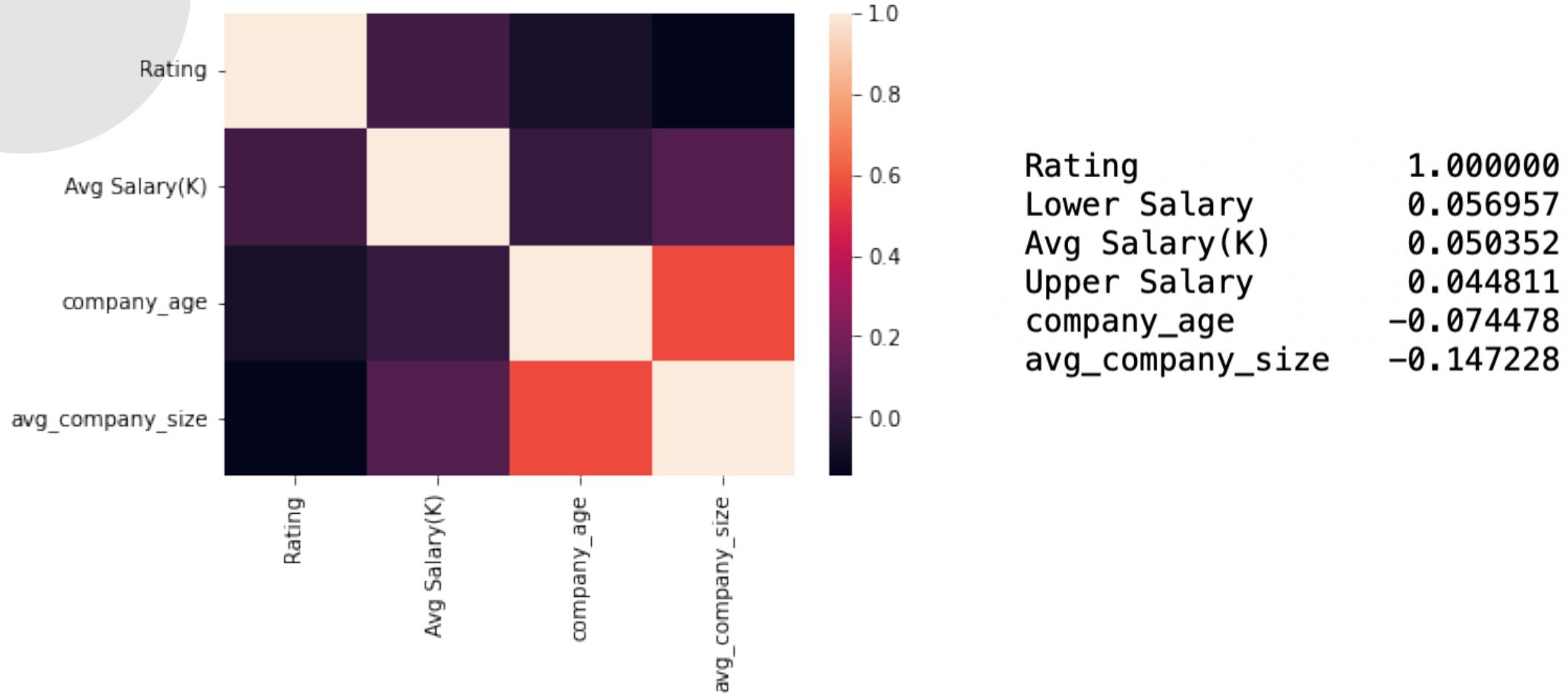
Results :

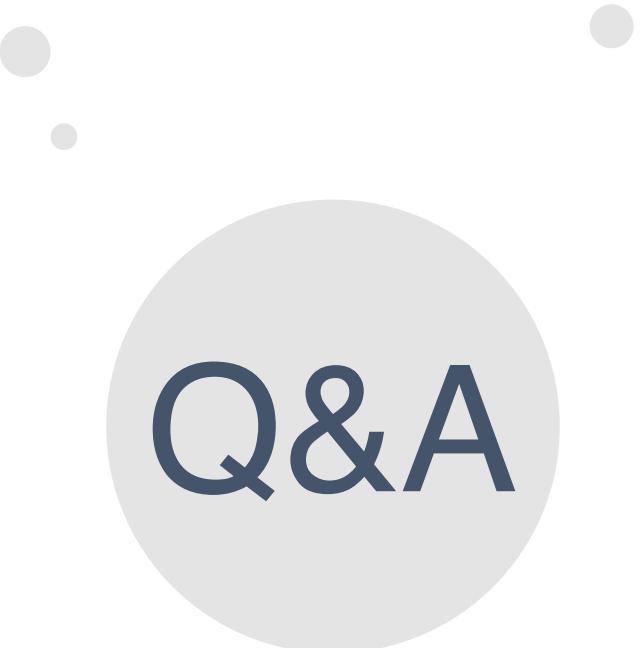
**1. Lower average salary level:
Healthcare, Food, Education**

**2. Higher average salary level:
Game, IT Services**

Hypothesis 4:

The rating of the company is somehow related to the company's age and size.





Q&A