



Data Science Salary Prediction

By Grace Gan



The Problem:

Lack of clear benchmarks, data scientists and recruiters are hard to settle down competitive compensation packages.

The Data

	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	Type of ownership	Industry	Sector	Revenue	Competitors
0	Data Scientist	\$53K-\$91K (Glassdoor est.)	Data Scientist Location: Albuquerque, NM	3.8	Tecolote Research	Albuquerque, NM	Goleta, CA	501 to 1000 employees	1973	Company - Private	Aerospace & Defense	Aerospace & Defense	\$50 to \$100 million (USD)	-1
1	Healthcare Data Scientist	\$63K-\$112K (Glassdoor est.)	What You Will Do:	3.4	University of Maryland Medical System	Linthicum, MD	Baltimore, MD	10000+ employees	1984	Other Organization	Health Care Services &	Health Care	\$2 to \$5 billion (USD)	-1
2	Data Scientist	\$80K-\$90K (Glassdoor est.)	KnowBe4, Inc. is a high growth information security company.	4.8	KnowBe4	Clearwater, FL	Clearwater, FL	501 to 1000 employees	2010	Company - Private	Security Services	Business Services	\$100 to \$500 million (USD)	-1
3	Data Scientist	\$56K-\$97K (Glassdoor est.)	*Organization and Job ID** Job ID: 310709	3.8	PNNL	Richland, WA	Richland, WA	1001 to 5000 employees	1965	Government	Energy	Oil, Gas, Energy & Business	\$500 million to \$1 billion (USD)	Oak Ridge National Commerce
4	Data Scientist	\$86K-\$143K (Glassdoor est.)	Data Scientist Affinity Solutions / Marketing	2.9	Affinity Solutions	New York, NY	New York, NY	51 to 200 employees	1998	Company - Private	Advertising & Marketing	Business Services	Unknown / Non-Applicable	Signals, Digital Realty,
5	Data Scientist	\$71K-\$119K (Glassdoor est.)	CyrusOne is seeking a talented Data Scientist who holds a range Job Description	3.4	CyrusOne	Dallas, TX	Dallas, TX	201 to 500 employees	2000	Company - Public	Real Estate	Real Estate	\$1 to \$2 billion (USD)	
6	Data Scientist	\$54K-\$93K (Glassdoor est.)		4.1	ClearOne Advantage	Baltimore, MD	Baltimore, MD	501 to 1000 employees	2008	Company - Private	Banks & Credit Unions	Finance	Unknown / Non-Applicable	-1
7	Data Scientist	\$86K-\$142K (Glassdoor est.)	Advanced Analytics – Lead Data Scientist SUMMARY	3.8	Logic20/20	San Jose, CA	Seattle, WA	201 to 500 employees	2005	Company - Private	Consulting	Business Services	\$25 to \$50 million (USD)	-1
8	Research Scientist	\$38K-\$84K (Glassdoor est.)		3.3	Rochester Regional Health	Rochester, NY	Rochester, NY	10000+ employees	2014	Hospital	Health Care Services &	Health Care	\$500 million to \$1 billion (USD)	-1
9	Data Scientist	\$120K-\$160K (Glassdoor est.)	isn't your usual company. Our work is powered by the premise At Wish, our Data Science & Engineering team is comprised of	4.6	<intent> Wish	New York, NY	New York, NY	51 to 200 employees	2009	Company - Private	Internet	Information Technology	\$100 to \$500 million (USD)	Clicktripz, SmarterTr
10	Data Scientist	\$126K-\$201K (Glassdoor est.)	Secure our Nation, Ignite your Future	3.5	ManTech	San Jose, CA	San Francisco, CA	501 to 1000 employees	2011	Company - Private	Other Retail Stores	Retail	\$1 to \$2 billion (USD)	-1
11	Data Scientist	\$64K-\$106K (Glassdoor est.)	Position Summary... Drives the execution of multiple	4.1	ManTech	Chantilly, VA	Herndon, VA	5001 to 10000 employees	1968	Company - Public	Research & Development	Business Services	\$1 to \$2 billion (USD)	-1
12	Staff Data Scientist - Technology	\$106K-\$172K (Glassdoor est.)	Are you an experienced Data Analyst, skilled at providing Job Description	3.2	Walmart	Plano, TX	Bentonville, AR	10000+ employees	1962	Company - Public	Department, Clothing, & Advertising & Marketing	Retail	\$10+ billion (USD)	Target, Costco
13	Data Analyst	\$46K-\$85K (Glassdoor est.)		4.1	Yesler	Seattle, WA	Seattle, WA	201 to 500 employees	2012	Company - Private	Biotech & Pharmaceutical	Business Services	Unknown / Non-Applicable	-1
14	Data Scientist	\$83K-\$144K (Glassdoor est.)		3.7	Takeda Pharmaceuticals	Cambridge, MA	OSAKA, Japan	10000+ employees	1781	Company - Public	Biotech & Pharmaceutical	Biotech & Pharmaceutical	\$10+ billion (USD)	Novartis, Baxter,
15	Data Engineer I	\$102K-\$190K (Glassdoor est.)	This opportunity is within Audibles Data Engineering group. Scientist I/II, Biology	3.6	Audible	Newark, NJ	Newark, NJ	1001 to 5000 employees	1995	Subsidiary or Business Segment	Motion Picture Production &	Media	Unknown / Non-Applicable	-1
16	Scientist I/II, Biology	\$67K-\$137K (Glassdoor est.)		3.9	Blueprint Medicines	Cambridge, MA	Cambridge, MA	51 to 200 employees	2011	Company - Public	Biotech & Pharmaceutical	Biotech & Pharmaceutical	\$1 to \$5 million (USD)	bluebird bio, Agios
17	Customer Data Scientist	\$118K-\$189K (Glassdoor est.)	Company Overview	4.3	h2o.ai	Mountain View, CA	Mountain View, CA	201 to 500 employees	2011	Company - Private	Enterprise Software &	Information Technology	Unknown / Non-Applicable	-1
18	Data Scientist - Health Data	\$110K-\$175K (Glassdoor est.)	At Nuna, our mission is to make high-quality healthcare	4.2	Nuna	San Francisco, CA	San Francisco, CA	51 to 200 employees	2010	Company - Private	Enterprise Software &	Information Technology	Unknown / Non-Applicable	-1
19	Data Scientist	\$64K-\$111K (Glassdoor est.)	Pinnacol Assurance does just one thing, and does it better than	4.0	Pinnacol Assurance	Denver, CO	Denver, CO	501 to 1000 employees	1915	Nonprofit Organization	Insurance Carriers	Insurance	\$500 million to \$1 billion (USD)	-1

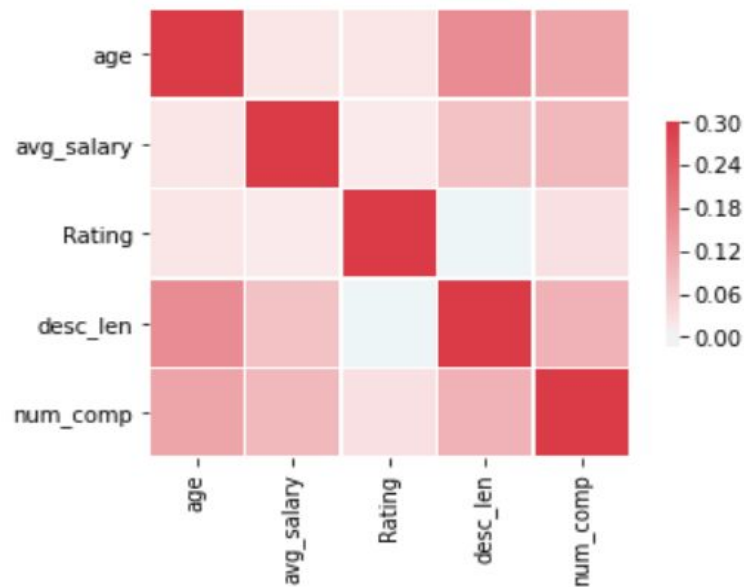
Data Wrangling

- **Parsing Numeric Data:** Extracted numeric values from salary estimates to create a usable salary column.
- **Employer-Provided Salary and Hourly Wages:** Created separate columns for employer-provided salaries and hourly wages.
- **Removing Incomplete Rows:** Removed rows that did not contain salary information to ensure a complete dataset.
- **Parsing Company Ratings:** Extracted the company rating from the text.
- **Company State:** Added a column for the state in which the company is located.
- **Headquarters Indicator:** Added a column indicating whether the job was located at the company's headquarters.
- **Company Age:** Transformed the company's founded date into a numerical column representing the age of the company.
- **Skills Columns:** Created columns indicating whether specific skills (Python, R, Excel, AWS, Spark) were mentioned in the job description.

Exploratory Data Analysis

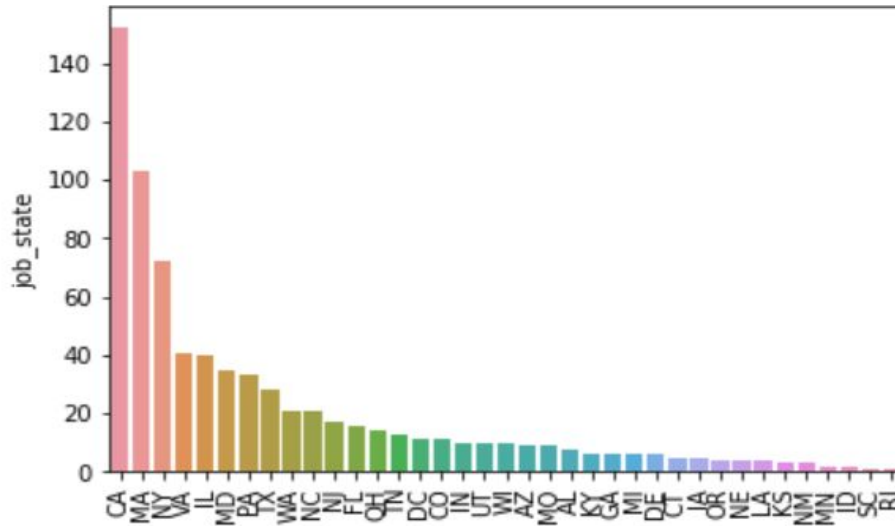
(1) 'age','avg_salary','Rating','desc_len' correlation?

	age	avg_salary	Rating	desc_len
age	1.000000	0.019655	0.021655	0.163911
avg_salary	0.019655	1.000000	0.013492	0.078808
Rating	0.021655	0.013492	1.000000	-0.012281
desc_len	0.163911	0.078808	-0.012281	1.000000

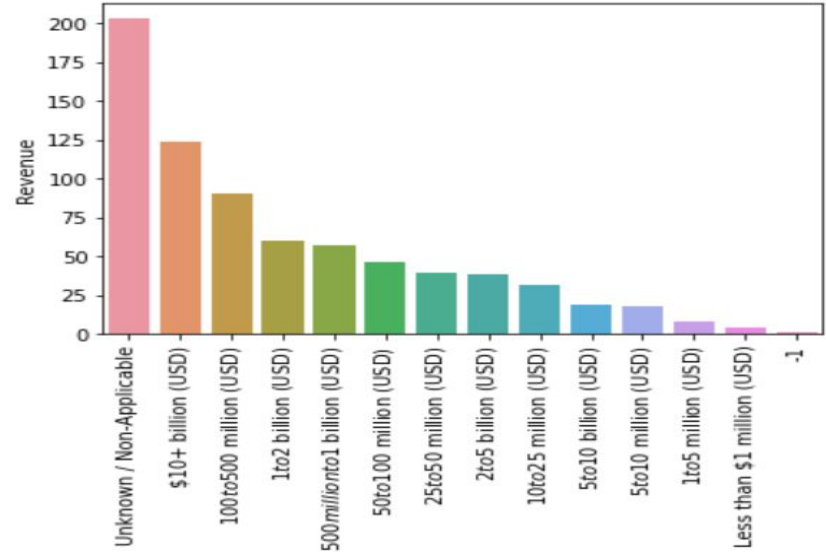


(2) Each categorical variable correlation?

graph for job_state: total = 37



graph for Revenue: total = 14



Model Selection

- ❖ **Multiple Linear Regression**
- ❖ **Lasso Regression**
- ❖ **Random Forest Regressor**

Conclusion

- ❖ **Random Forest Regressor: MAE = 11.22**
- ❖ **Linear Regression: MAE = 18.86**
- ❖ **Ridge Regression: MAE = 19.67**

Conclusion

- The **Random Forest** model far outperformed the other approaches
- For job seekers, HR professionals, or hiring managers, using this model to inform salary negotiations, an MAE of **\$11,220** signifies the typical error margin. This helps in setting realistic salary expectations and negotiating offers. For example, if the model predicts a salary of \$100,000, the actual salary could typically be expected to fall within the range of \$88,780 to \$111,220.

Thank You!