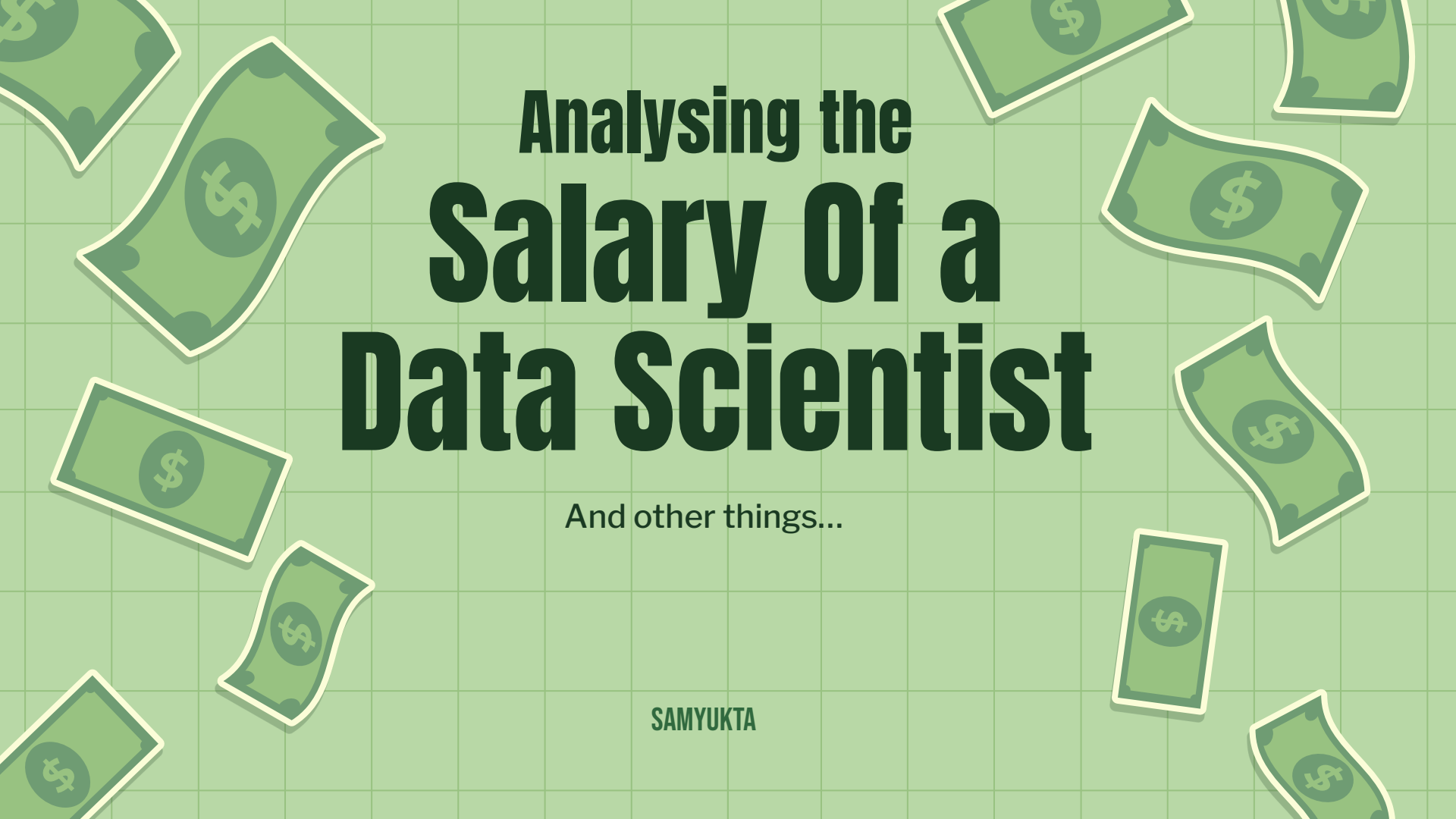


Analysing the Salary Of a Data Scientist

And other things...

SAMYUKTA

Contents

Intro to the Data

This was taken from Kaggle, which again was extracted from glassdoor.
We have chosen the following columns from 42 columns:

1. **Job_Title** (Data Scientist, Data Analyst..ect)
2. **Company_Name**
3. **Rating (for the company)** (From 1-5)
4. **Location**
5. **Size of company**
6. **Sector**
7. **Minimum Salary and Maximum Salary**
8. **Age of company**
9. **Language_used** (Python, Excel, SQL, Power BI..ect)

Part 1

APPLIED STEPS

- Source
- Promoted Headers
- Removed Columns
- Reordered Columns
- Removed Columns1
- Reordered Columns1
- Replaced Value
- Replaced Value1
- Filtered Rows
- Removed Columns2
- Renamed Columns

Replace Values

Replace one value with another in the selected

Value To Find

Company - Private

Replace With

Private

Filter Rows

Apply one or more filter conditions to the rows in this table.

Basic Advanced

Keep rows where 'Sector'

does not equal -1

And Or

Enter or select a value

Part 2 - Creating the Dimension Tables

Job_Title_ID	Job_Title
1	data scientist
2	other scientist
3	analyst
4	data engineer
5	data analitics
6	na
7	data modeler
8	Data scientist project …
9	machine learning engine…
10	director

- Job Title Dim
- Company Dim
- Location Dim
- Type of ownership Dim
- Sector Dim
- Seniority_by_title Dim
- Degree Dim
- Fact Table Final

Company_ID	Company
1	Tecolote Research
2	University of Maryland Medic…
3	KnowBe4
4	PNNL
5	Affinity Solutions
6	CyrusOne
7	ClearOne Advantage
8	Logic20/20
9	Rochester Regional Health

Ownership_ID	Type of ownership
1	Private
2	Other Organization
3	Government
4	Public
5	Hospital
6	Subsidiary or Business …
7	Nonprofit Organization
8	College / University
9	School / School District

	Sector_ID	Sector
1	1	Aerospace & Defense
2	2	Health Care
3	3	Business Services
4	4	Oil, Gas, Energy & Utiliti…
5	5	Real Estate
6	6	Finance
7	7	Information Technology
8	8	Retail
9	9	Biotech & Pharmaceuticals
10	10	Media
11	11	Insurance
12	12	Transportation & Logistics

Part 3 - Using Split

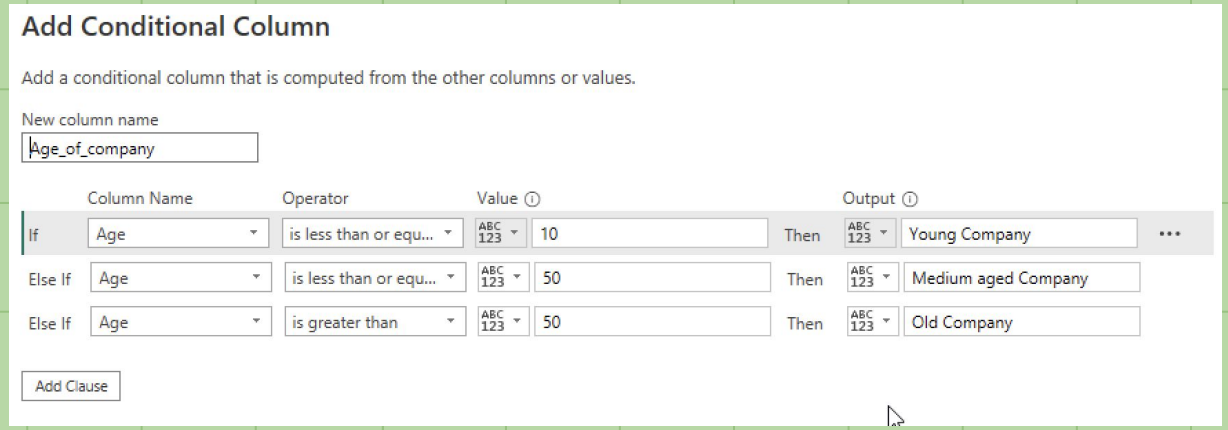
Loc_ID	Location
1	Albuquerque, NM
2	Linthicum, MD
3	Clearwater, FL
4	Richland, WA
5	New York, NY
6	Dallas, TX
7	Baltimore, MD
8	San Jose, CA
9	Rochester, NY
10	Chantilly, VA
11	Plano, TX
12	Seattle, WA
13	Cambridge, MA
14	Newark, NJ
15	Mountain View, CA
16	San Francisco, CA
17	Denver, CO
18	Chicago, IL
19	Louisville, KY
20	Herndon, VA
21	Hillsboro, OR
22	Worcester, MA
23	Groton, CT
24	Detroit, MI

Loc_ID	Location.1	Location.2
1	Albuquerque	NM
2	Linthicum	MD
3	Clearwater	FL
4	Richland	WA
5	New York	NY
6	Dallas	TX
7	Baltimore	MD
8	San Jose	CA
9	Rochester	NY
10	Chantilly	VA
11	Plano	TX
12	Seattle	WA
13	Cambridge	MA
14	Newark	NJ
15	Mountain View	CA
16	San Francisco	CA
17	Denver	CO
18	Chicago	IL
19	Louisville	KY
20	Herndon	VA
21	Hillsboro	OR
22	Worcester	MA
23	Groton	CT
24	Detroit	MI

Conditional Company Age

Age
48
151
37
11
11
11
11
56
56

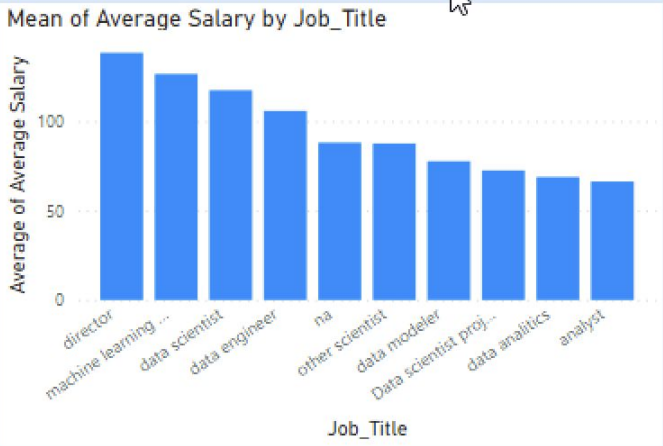
Age_of_company
Medium aged Company
Old Company
Medium aged Company
Medium aged Company
Medium aged Company
Medium aged Company
Medium aged Company
Old Company
Old Company
Old Company

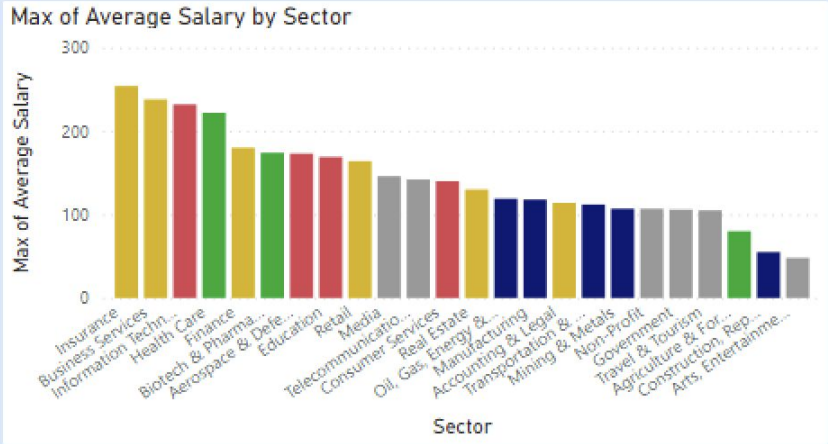


Final Relationship


Time for
Visualisation

Data Scientist Job Analysis



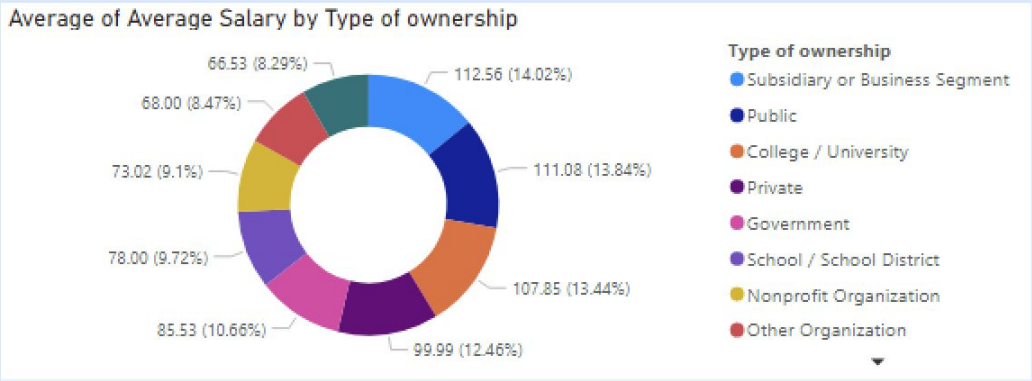


127.89

Upper_Sal_AVG

74.02

Lower_Salary_AVG



Data Scientist Job Analysis

Python

Sum of Average Salary
80749

Tableau

Degree	analyst	data analitics	data engineer	data modeler	data scientist	Data scientist project manager	director	machine learning engineer	na	other scientist	Total
M	12	3	4	0	33	2		0	0	0	**54**
P			0		0		0	0		0	**0**
Unobserved	30	1	8	0	46	10	0	0	6	0	**101**
Total	**42**	**4**	**12**	**0**	**79**	**12**	**0**	**0**	**6**	**0**	**155**

SQL

Degree	analyst	data analitics	data engineer	data modeler	data scientist	Data scientist project manager	director	machine learning engineer	na	other scientist	Total
M	27	3	28	3	85	1		2	0	0	**149**
P			2		14		0	4		0	**20**
Unobserved	54	1	64	1	86	10	0	7	8	4	**235**
Total	**81**	**4**	**94**	**4**	**185**	**11**	**0**	**13**	**8**	**4**	**404**

Sum of Upper Salary and Count of Loc_ID by Age_of_company and Job_Title

96