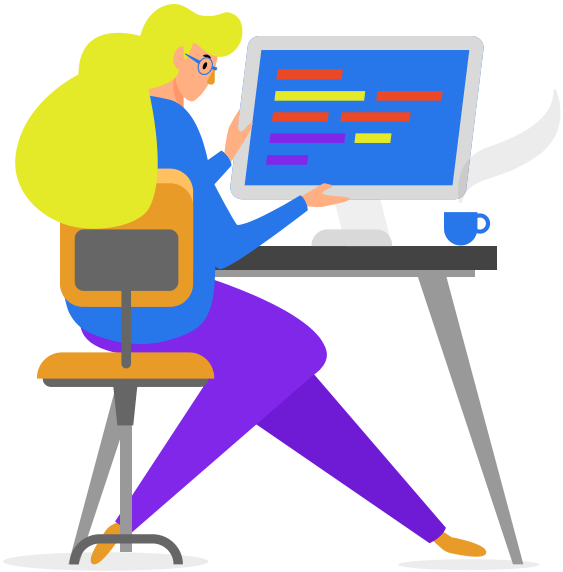
Scrapping & Analysis of Job Portals

Process



01

Web Scraping

Indeed and Glassdoor were scrapped.

02

Job Categories

Data Scientist, Big Data Analyst and Machine Learning were the categories that were scrapped.

03

Data Cleaning

Over 5000+ raw data were scrapped which were later cleaned.

04

Model Learning

Different Regression Models were tried to find the correct fit.

05

Model Evaluation

Models were evaluated using performative metrics

06

Integrate the model

Dashboard was created to highlight the predictions

Web Scraping

BeautifulSoup

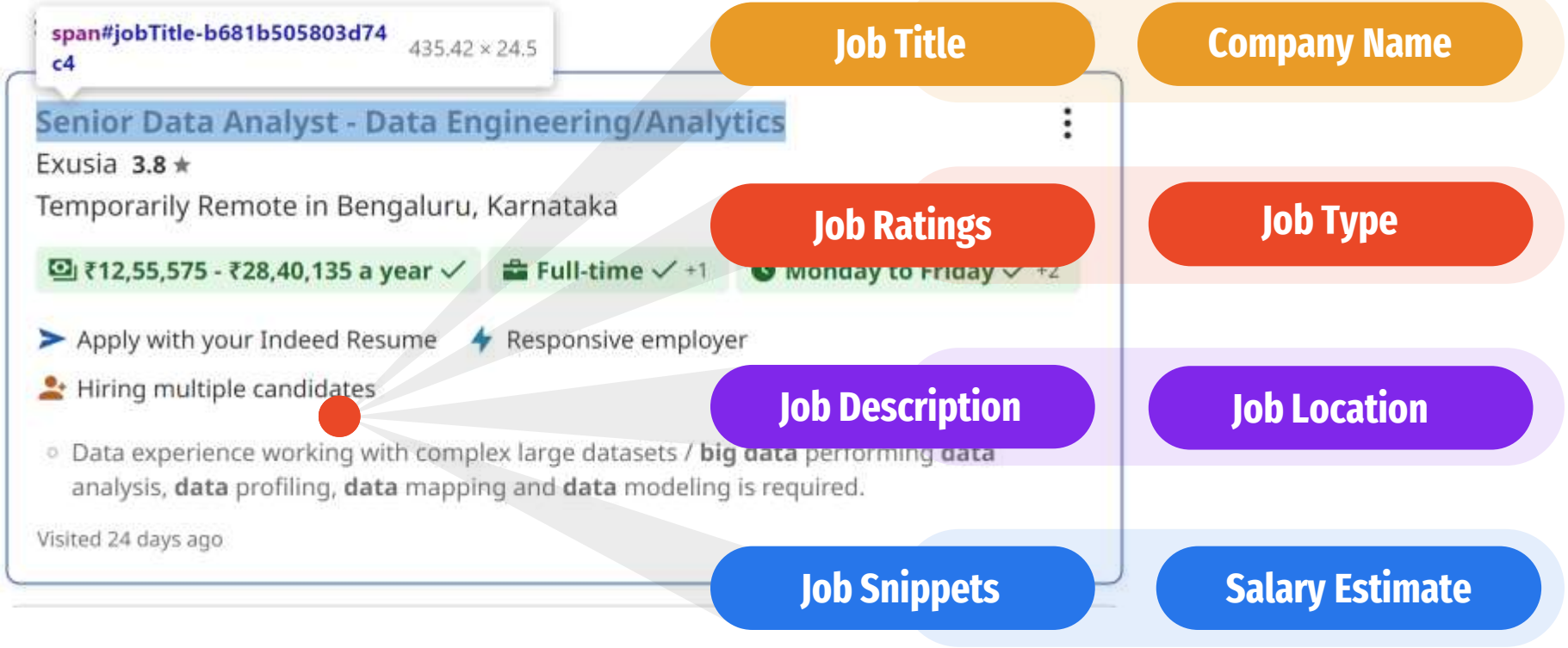
- Beautiful Soup is a very powerful library that makes web scraping by traversing the DOM easier to implement.
- It does only static scraping & ignores JavaScript. It fetches web pages from the server without the help of a browser. You get exactly what you see in *"view page source"*, and then you slice and dice it.

Vs

Selenium

- Selenium automates web browser interaction from python.
- The combination of Beautiful Soup and Selenium will do the job of dynamic scraping.
- The data rendered by JavaScript links can be made available by automating the button clicks with Selenium and then can be extracted by Beautiful Soup.

Web Scraping

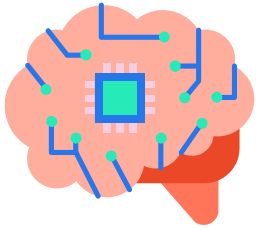


Data Cleaning

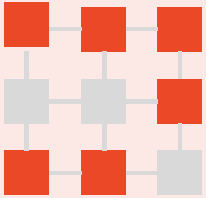
01

Raw Data

- Jobs were categorized in 11 different sub category.
- Jobs were classified into three levels: Entry , Middle and Advance Level.



02



Cleaned Data

- Years of Experience was scraped from Job Description
- Job Type was levelled into : Full Time, Part Time and Night Shift.

Missing Data

Machine learning

Imputations were done after learning the trends among the data variables.

Location

Location containing India/Remote were removed.

04

Ratings

Ratings were imputed using the mean.

03

Job Titles

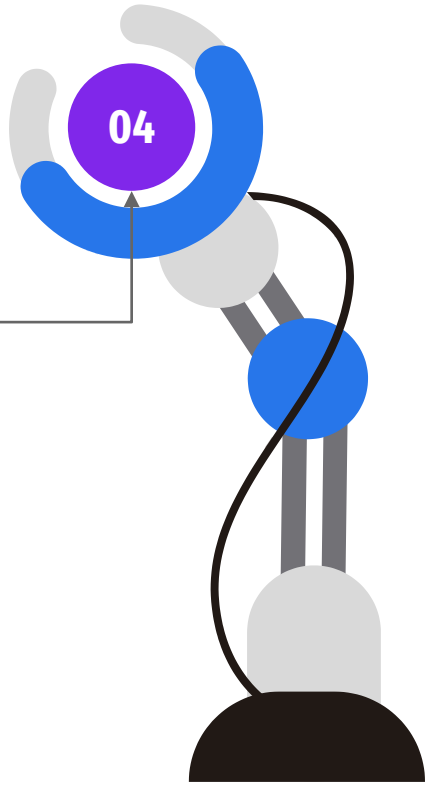
All NA Job Titles were removed

02

Years of Experience

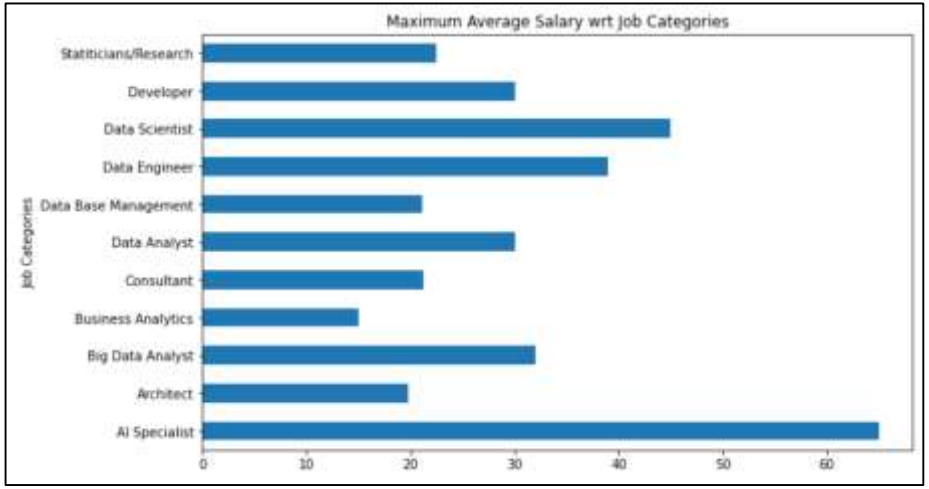
Random Number Generator:
Entry : 0-1
Mid : 1-5
Advance:5-10

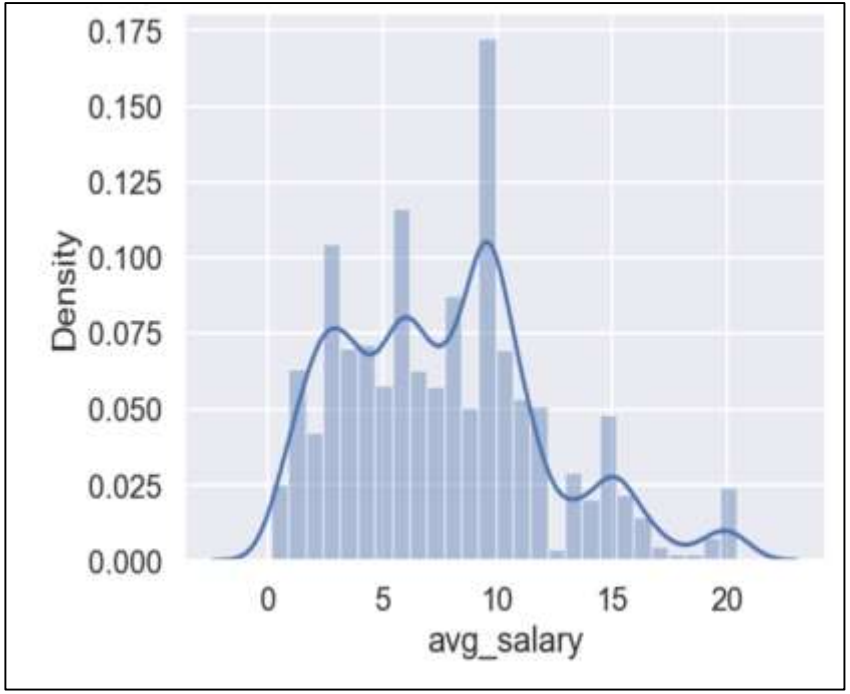
01



Data Analyst
Data Science
Analyst Data
Engineer Machine
Data Scientist
Learning Engineer
Machine Learning
Scientist Data
Business Analyst
Senior Data
Analyst Business
Python Machine
Research Analyst
Data Engineering
Operations Analyst
Intern Machine
Manager Data
Senior Analyst
Intern Data
Analyst Senior
Security Analyst
Artificial Intelligence
Lead
Analyst Machine
Analyst Product
Sr Data
Principal
Engineer Jr
Reporting Analyst
Learning Intern
Learning Data
Sr Machine
Developer
Learning Developer
Associate Data
Analyst BA
Master Data
Analytics
Junior Data
Science Senior
Financial Analyst
Researcher University
Innovation Center
Scientist Chase
Scientist Mentor
Doctoral Researcher
Science Intern
Developer Machine
Software Engineer
Visa Consulting

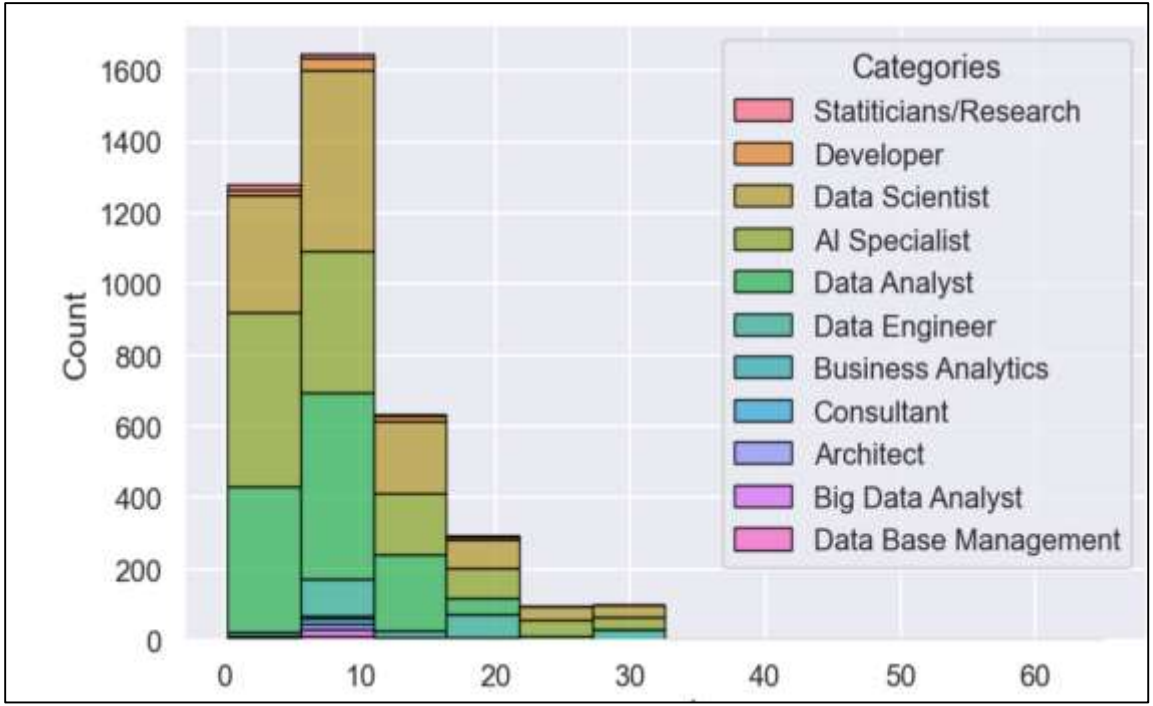
Statistical Analysis





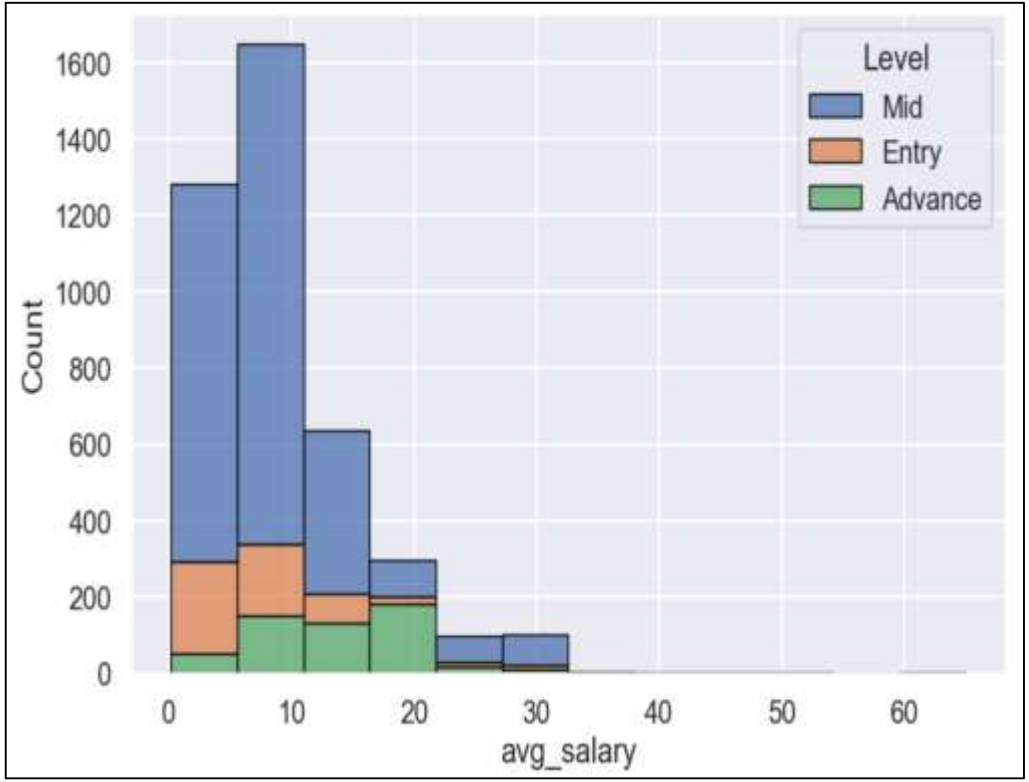
Statistical Analysis

Categories	avg_salary
Data Engineer	11.627546
Developer	9.537267
Architect	9.512343
Data Base Management	9.512343
Data Scientist	9.500000
Consultant	8.892957
Big Data Analyst	8.857252
Business Analytics	8.103528
AI Specialist	7.500000
Data Analyst	7.000000
Statiticians/Research	6.593858

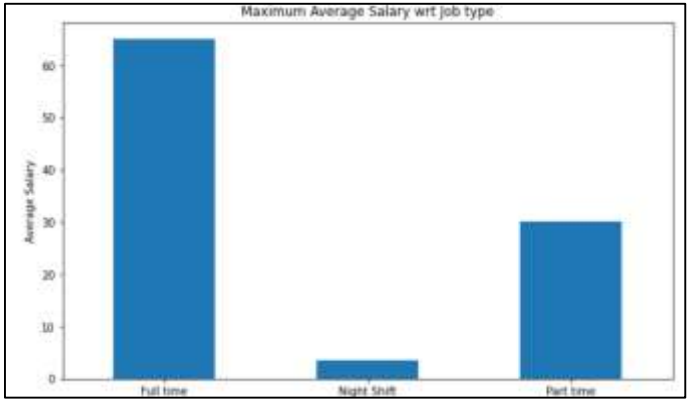


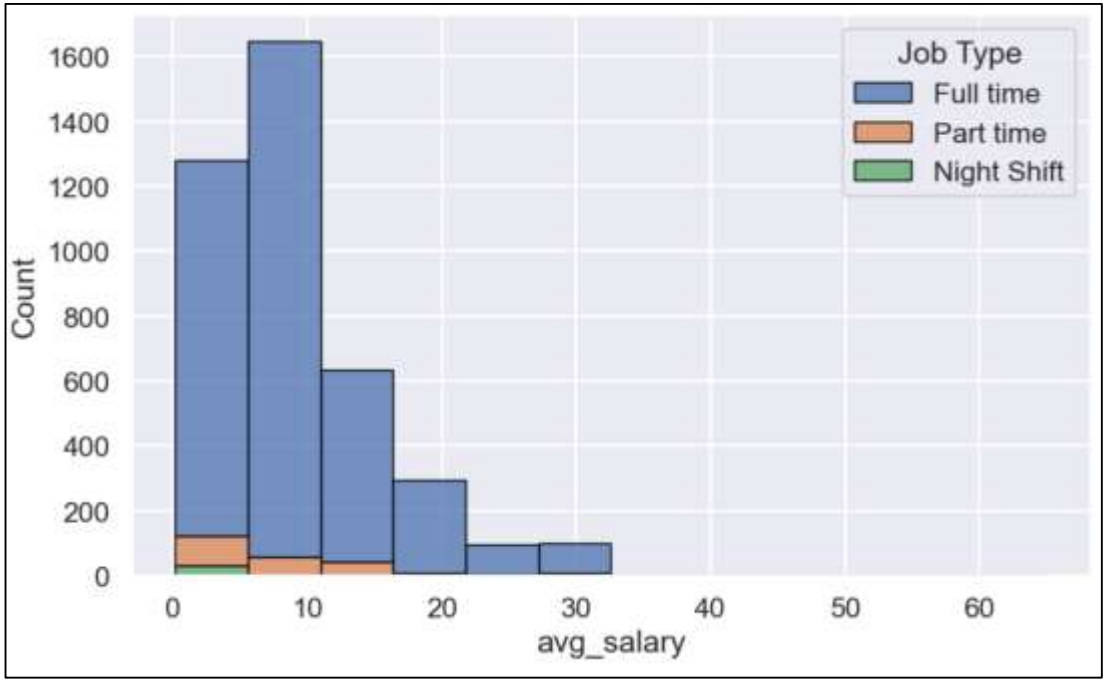
Statistical Analysis

Level	avg_salary
Advance	15.160239
Entry	5.770436
Mid	8.000000



Statistical Analysis



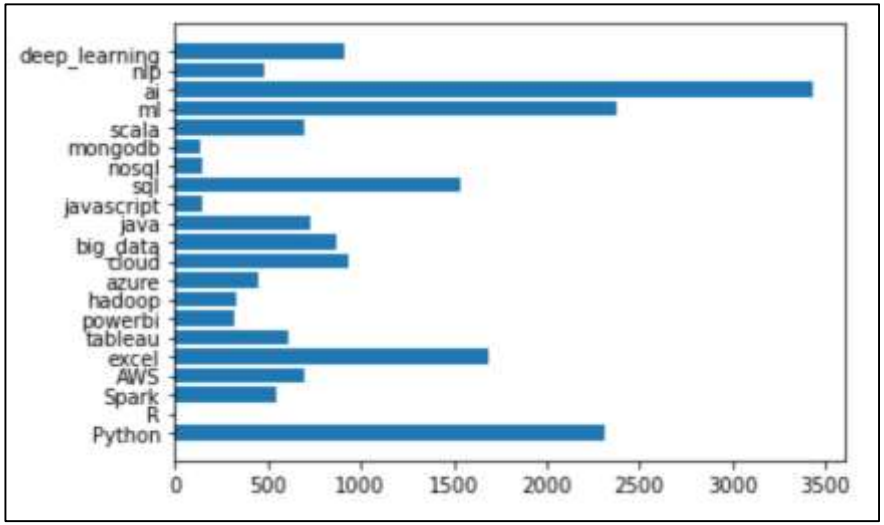


Statistical Analysis

Categories	Level	avg_salary
AI Specialist	Advance	11.500000
	Entry	5.594284
	Mid	7.500000
Architect	Advance	13.613761
	Mid	8.755332
Big Data Analyst	Advance	12.249238
	Entry	6.105507
	Mid	9.184797
Business Analytics	Advance	10.205235
	Entry	8.103528
	Mid	7.099763
Consultant	Advance	14.770855
	Entry	7.131056
	Mid	9.512343
Data Analyst	Advance	15.045758
	Entry	3.360000
	Mid	6.500000

Data Base Management	Advance	15.570311
	Entry	4.350741
	Mid	9.512343
Data Engineer	Advance	19.749859
	Entry	5.594284
	Mid	9.217112
Data Scientist	Advance	13.984339
	Entry	6.500000
	Mid	8.755332
Developer	Advance	15.791741
	Entry	5.770436
	Mid	9.537267
Statiticians/Research	Advance	15.498102
	Entry	8.000000
	Mid	5.220000

Statistical Analysis



Bangalore/Mumbai/Pune

Categories	avg_salary
Data Engineer	15.498102
Consultant	14.000000
Architect	9.537267
Data Base Management	9.512343
Data Scientist	9.512343
Developer	8.755332
Big Data Analyst	8.627666
AI Specialist	7.500000
Data Analyst	7.234361
Business Analytics	7.131056
Statiticians/Research	5.220000

Categories	avg_salary
Statiticians/Research	13.078679
Developer	11.518633
Consultant	9.763438
Big Data Analyst	9.512343
Data Engineer	8.273570
Data Scientist	8.273570
Data Analyst	6.500000
Business Analytics	6.227577
AI Specialist	5.594284

Categories	avg_salary
Business Analytics	10.205235
Data Engineer	9.537267
Big Data Analyst	9.512343
Data Scientist	9.500000
Architect	8.755332
Developer	8.431600
Data Base Management	8.273570
AI Specialist	7.500000
Consultant	7.131056
Data Analyst	6.367192

Link to Dashboard

Added in the Zip file.