

26 November 2023

ASSIGNMENT - JOB MARKET ANALYSIS

1. PROJECT AIM

Analysis of job market data to explore the open positions related to the jobs in the field of data or machine learning in a particular region. While it is not compulsory to do so, you could use web scraping techniques introduced in this course to collect data from these websites. Also, note that you are not limited to the sites mentioned above. Identity one job that you aspire to transition into and identify the skills required for the job.

A. Objectives & Assignment Description:

- i. Collect and use job market data to explore data related positions or machine learning related positions in a particular region (e.g., US, India).
- ii. You will have to use appropriate keywords (for e.g., “data scientist”, “data engineer”, “ML engineer”) to extract information on the positions available on websites of your choice.

2. PROCEDURES:

A. Data Collection Method

The data collection involves collecting the data from any one of the websites like Indeed, Glassdoor, Naukri, Monster. We also must use one keyword to extract the required jobs from website. For this assignment I choose the following.

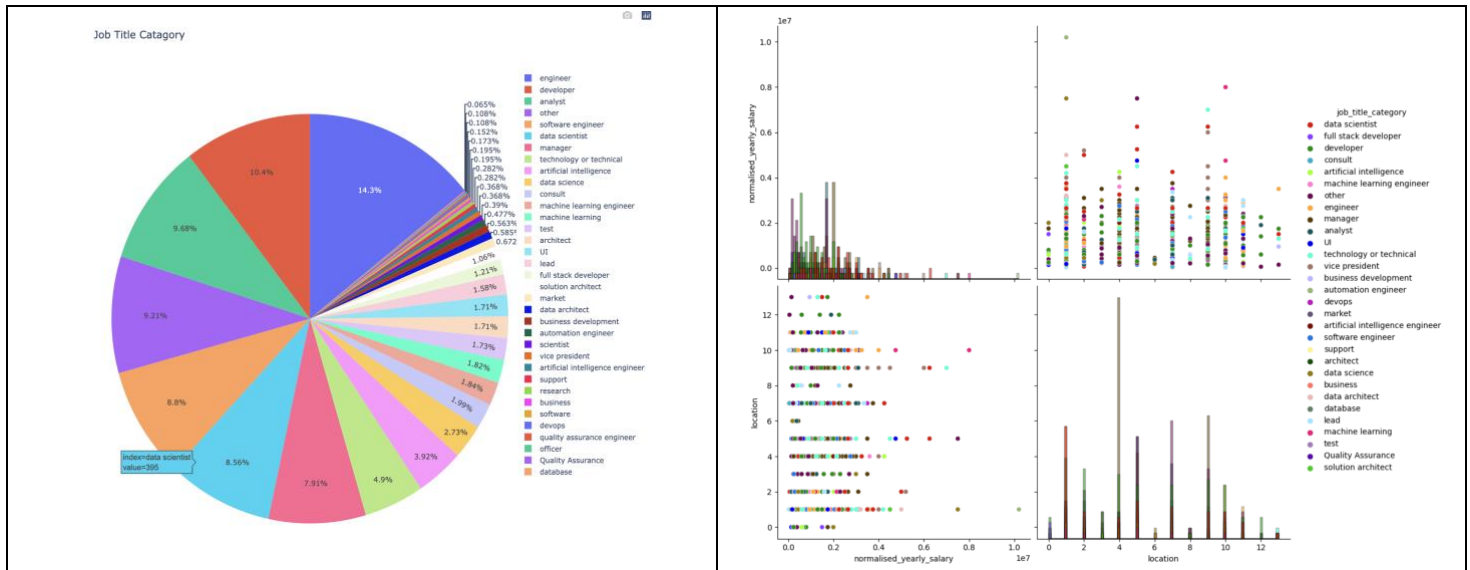
- Website – Indeed
 - Keyword – Machine Learning
- i. Robot text file from indeed does not block /jobs and /viewjob URLs for user agent *. We also inspected the page to find corresponding html tags, classes and other attributes to extract the data.
 - ii. **Data discovery:** I have selected the indeed website for data extraction. Challenges faced are the following. The website uses Cloudflare for protection against DDoS mitigation, a Web Application Firewall, API protection, etc. But this also make sure we do not get the data by using simple requests library. I have also tries using special headers. But this was also not successful. Hence, I have

v. Enriching Data: Enriching the data is done by using the using nltk to extract major programming language tokens. We also did the same for job type from metadata. We have full-time, contract, fresher, internship, part-time, temporary, permanent, freelance as job types and these categories can be overlapping. We also extracted salary from metadata and separate it into maximum, minimum and normalized salary along with intended. Here normalized salary means salary calculated for entire year. We also had to create category for jobs as job title can be very similar but different when be compare. We have identified major clusters using K-means for the title and used this as a reference to create our major category of jobs.

[illegible]

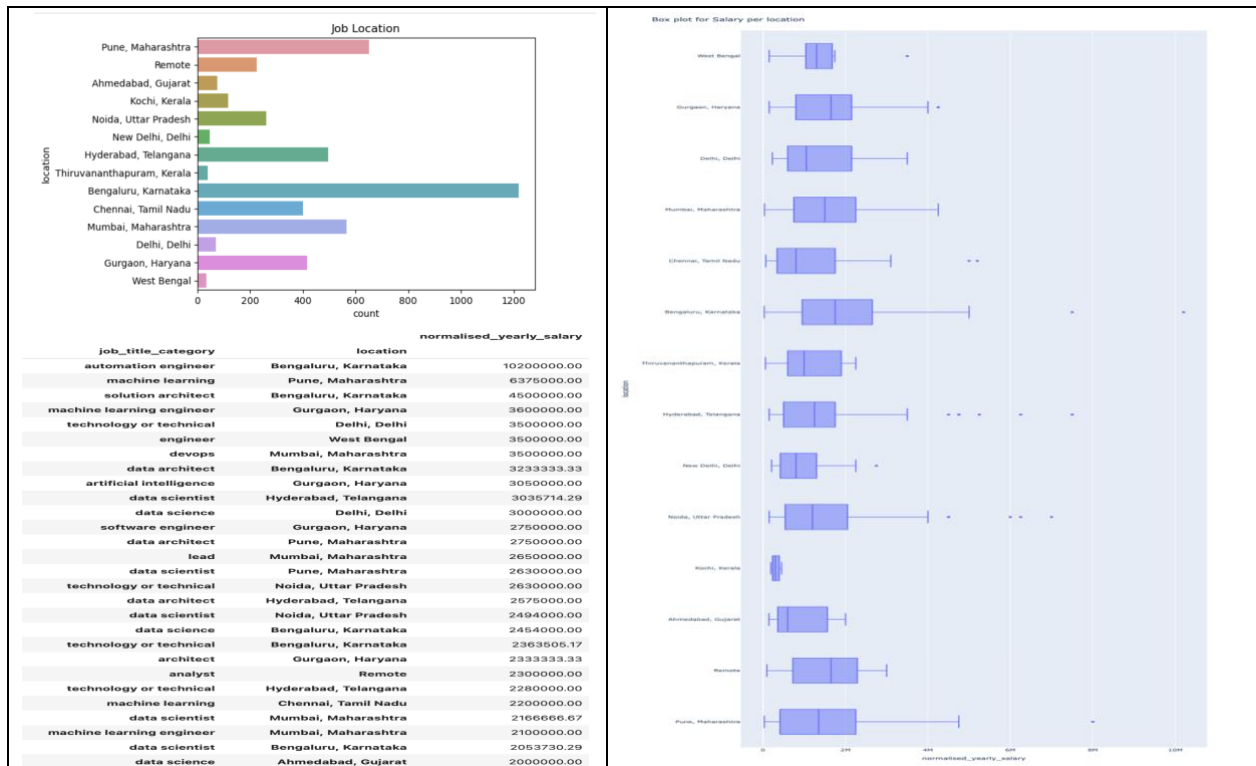
B. Market Data Visualization

i. Titles



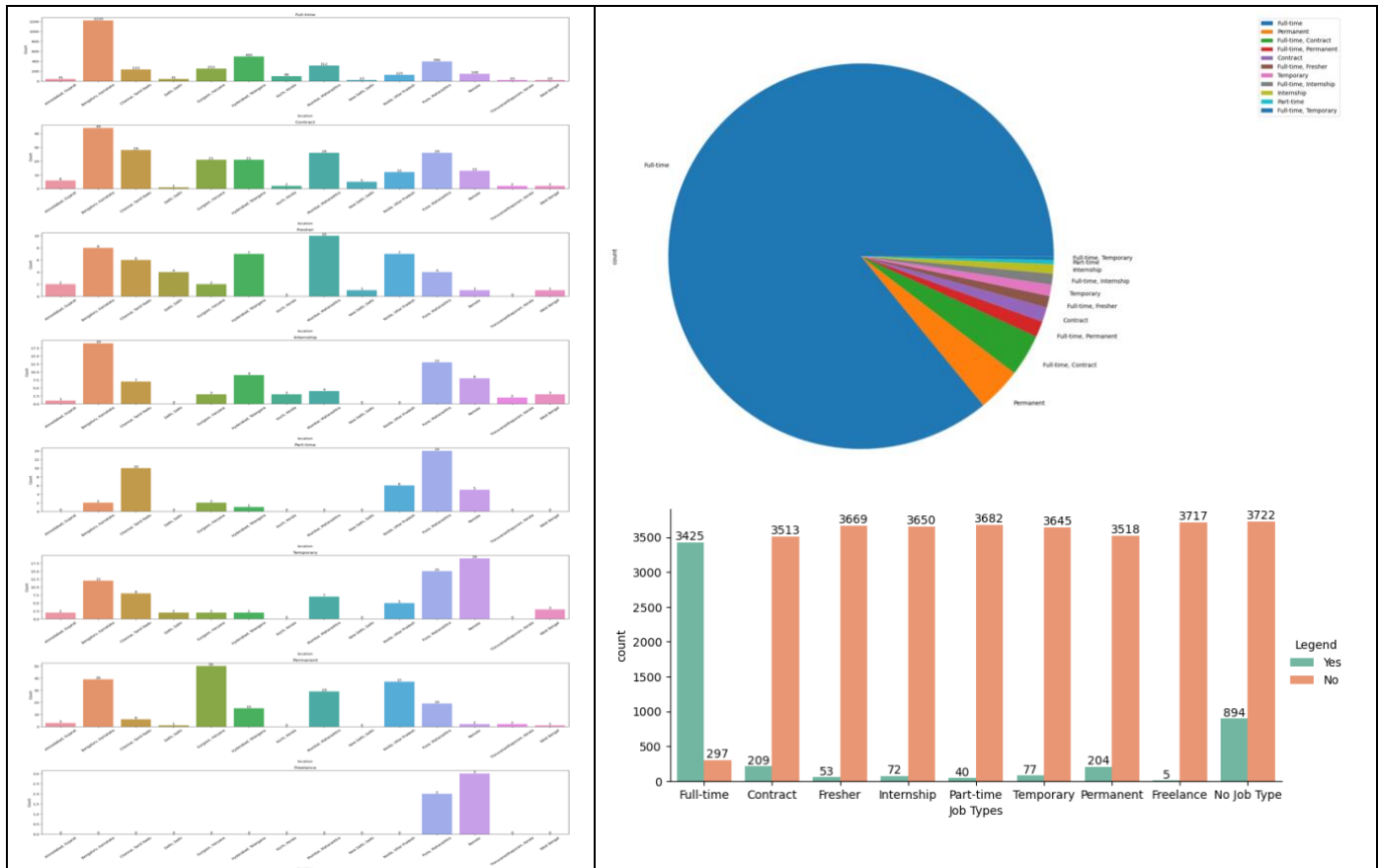
We can see that top categories are engineer, developer, analyst, software engineer, data scientist, manager, technology related, AI, Data science, consult, machine learning etc. In comparison to salary, other than positions like management and leadership, most of the jobs are well distributed in the category. But we can see some jobs in data scientist, machine learning and artificial intelligence as the highest salary job.

ii. Locations where jobs are offered.



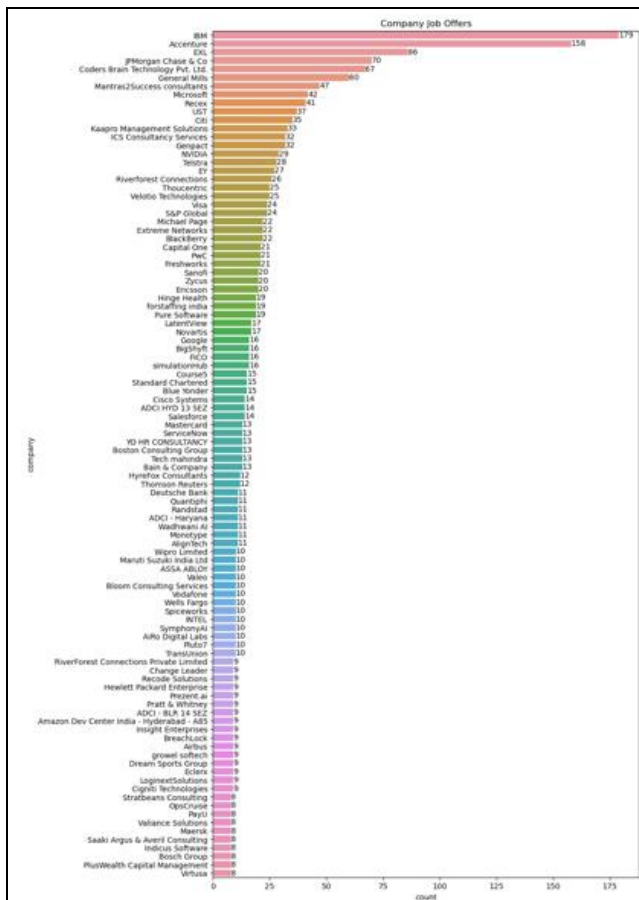
We can see most jobs have demand in Bangalore, Pune, Mumbai and Hyderabad. We can also see salary also follows the same order.

iii. Job Types



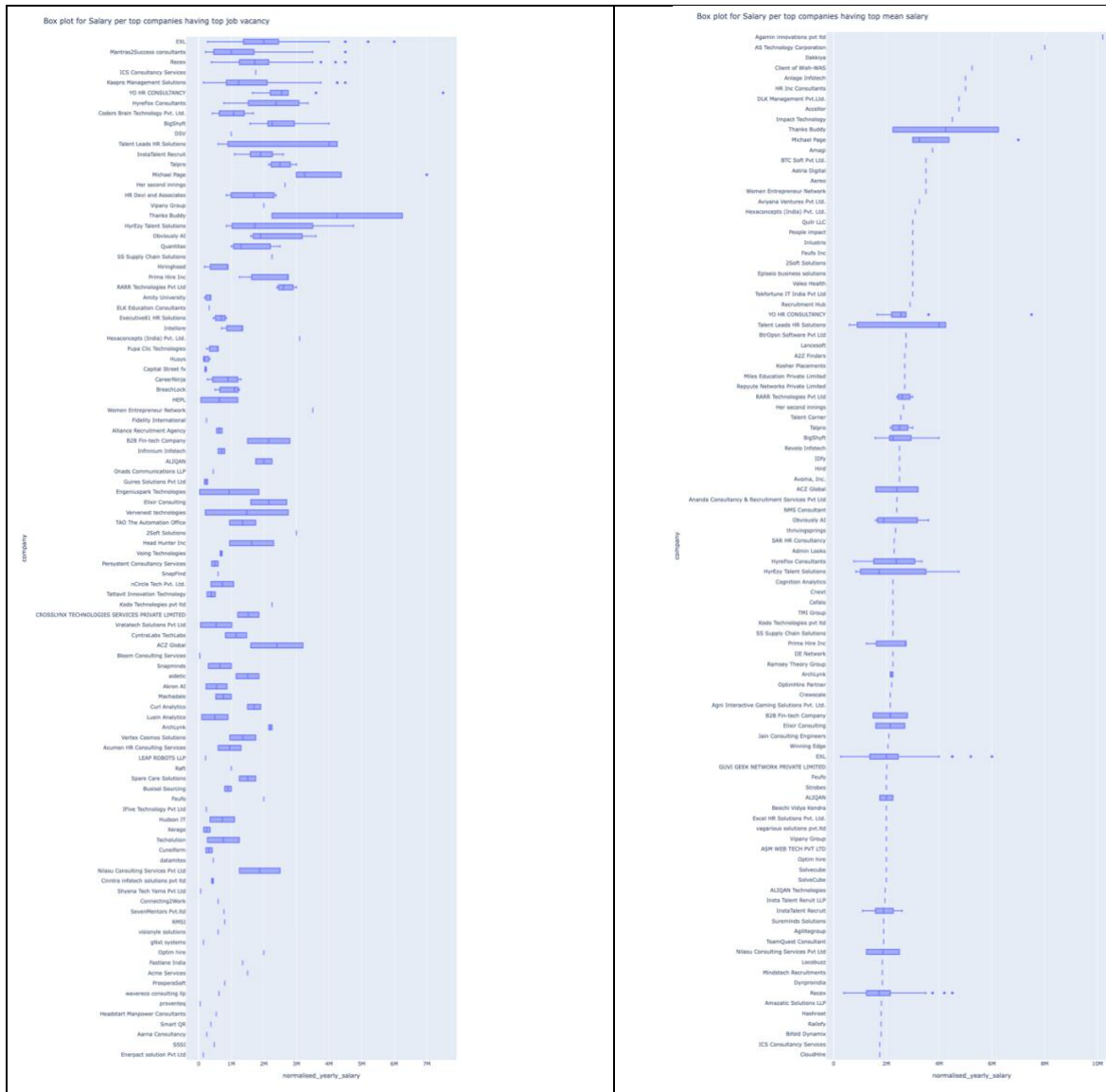
Note that pie chart only contains the top combination of the job types. We can see that full time employee is having highest demand. Part time is only having 10th position in demand.

iv. Company Offers.

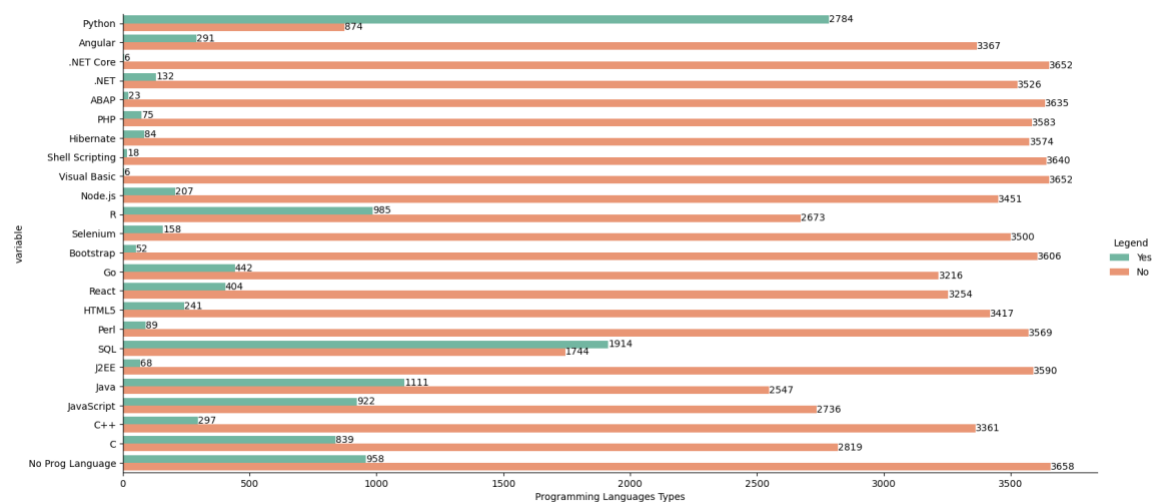


In the dataset, a significant number of companies are observed to have a limited job offering, often less than 10 positions. This trend persists as we narrow down our focus to the top 100 companies. It is noteworthy that there is a total of 1549 companies in the entire dataset out of which thousands of companies have only very few openings.

Furthermore, the salary distribution across different companies indicates a variation in compensation. Notably, only a few companies offer the highest-paying positions, and these tend to be specific roles with limited availability.



v. Skills



The top demanded skills are Python, SQL, Java, R, JavaScript, C, Go, React etc.

vi. Rating

	job_title	company	location	date	job_desc	metadata	rating	href	Full-time	Python
2240	ML Engineer III	Turing	Bengaluru, Karnataka	Posted 13 days ago	Solid understanding of machine learning algorithms	Full-time	4.80	https://in.indeed.com/rc/clk?jk=63ee2469b824ee...	True	True
3173	Senior Business Analyst	CoinDCX	Bengaluru, Karnataka	Posted 24 days ago	Build and deploy machine learning models where...	Full-time	4.80	https://in.indeed.com/rc/clk?jk=df76e526b80126...	True	True
1184	Data Scientist	Curl Analytics	Bengaluru, Karnataka	EmployerActive 5 days ago	In this role, you will be responsible for desi...	Up to ₹19,03,336 a year Full-time +1 Monday to...	4.70	https://in.indeed.com/company/Curl-Analytics/j...	True	True
1676	Full Stack Developer	Curl Analytics	Bengaluru, Karnataka	Posted 10 days ago	Curl works on developing Products & Solutions ...	₹5,00,000 - ₹25,00,000 a year Full-time	4.70	https://in.indeed.com/rc/clk?jk=9040bc7c84c209...	True	True
1097	Data Science - AI/ML with Big data Architect	Anlage Infotech	Bengaluru, Karnataka	EmployerActive 29 days ago	As a Solution Architect, the candidate will be...	Up to ₹50,00,000 a year Full-time +1	4.70	https://in.indeed.com/company/Anlage-Infotech/...	True	True
1337	Data Scientist – Technology	Anlage Infotech	Bengaluru, Karnataka	Posted 30+ days ago	Expertise in at least one of the following fie...	Full-time	4.70	https://in.indeed.com/rc/clk?jk=60dd8833eb98d7...	True	True
2933	RPA Developer (Automation Anywhere)	TAO The Automation Office	Bengaluru, Karnataka	EmployerActive 3 days ago	As an end-to-end automation solutions partner,...	₹7,00,000 - ₹12,00,000 a year Full-time +1 Mon...	4.50	https://in.indeed.com/company/TAO-The-Automati...	True	True
2934	RPA Developer (UiPath)	TAO The Automation Office	Bengaluru, Karnataka	EmployerActive 3 days ago	Experience – 3 to 5 Years*\nWe are looking fo...	₹15,00,000 - ₹20,00,000 a year Full-time Monda...	4.50	https://in.indeed.com/company/TAO-The-Automati...	True	True
3433	Senior ML Engineer	SymphonyAI	Bengaluru, Karnataka	Posted 30+ days ago	SymphonyAI Summit is an AI-powered IT and Ente...	Full-time	4.50	https://in.indeed.com/rc/clk?jk=80f807307b7c48...	True	True
3287	Senior Data Scientist	Scienaptic Systems	Bengaluru, Karnataka	Posted 30+ days ago	As a Senior Data Scientist you will play a cri...	Full-time	4.50	https://in.indeed.com/rc/clk?jk=6841f99657bee5...	True	True
429	Azure BI Engineer(Azure+Power BI)	Infiniti Research	Bengaluru, Karnataka	EmployerActive 20 days ago	4+ years of experience in Power BI Visualizati...	Full-time Day shift +1	4.40	https://in.indeed.com/company/Infiniti-Researc...	True	True
3673	Senior Software Engineer – Backend	Digital Green	Bengaluru, Karnataka	Posted 30+ days ago	As a key member of the technology team, the po...	Full-time	4.40	https://in.indeed.com/rc/clk?jk=8c0f6d030bad8b...	True	True

We can also see that the Turing, CoinDCX, Curl Analytics, Anlage Infotech, TAO The Automation Office, SymphonyAI, Scienaptic Systems, Infiniti Research, Digital Green, Yara, Sophos, Salesforce, Google, Elixir Consulting, EXCELUS LEARNING SOLUTIONS PVT LTD, Bain & Company, Adobe, SAP, NVIDIA, NEC Software Solutions and Microsoft is having the highest ratings from other users. This might show the trust these company have with the public.

C. My Ideal Job

i. My ideal job has following preferences.

1. *High Salary*
2. *Good Company - High rating*
3. *Better Location*
4. *Scope for improvement.*
5. *Skills – Python, R, SQL, Go, Machine learning, etc.*

I will prefer the Bengaluru as the preferred location as this location shows highest demand in job and highest variance in the salary. This indicates the potential for having

highest promotions and salary hikes which shows scope of improvement in the job position. The same location also has relatively cool climate and I personally prefer this place as my relatives are also in same city. I will conclude that the ideal job is Software Engineer, Machine Learning, Google Assistant (ID:3958) from Google. But If I only consider the jobs for which salary is given, then I will choose Data Science - AI/ML with Big data Architect (ID: 1097) from Anlage Infotech company.