Course 1: Foundation of Information

Project Part - A: 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.

Objective:

This research aims to comprehensively analyse the data scientist job market in India, leveraging data extracted from the Naukri platform through web scraping. The primary objectives include exploring job titles, required skills, experience levels, and company specifications, with a focus on key indicators such as open positions, geographical distribution, salary variations, and emerging trends. The research seeks to offer valuable insights for individuals, employers, and stakeholders by providing a detailed view of the data science employment landscape in India.

Data Collection Method:

Naukri, a prominent Indian job portal, was scraped using Selenium for data collection. The dataset focuses on data scientist positions, extracting key details like job title, company, location, experience, skills, and salary. The scraped data was exported in CSV for visualization.

	A	B	C	D	E	F	G	H
1	Title	Company	Rating	rience Ne	Salary	Location	Description	Required Skills
2	Data Scientist/ Se	Fractal Analyti		4 4-9 Yrs	Not disclo	Mumbai, Pune, Chennai,	Demonstrate excellent communication skills and the ability to work effectively in a tea...	['Computer science', 'GIT', 'GCP', 'Analytical', 'Software development life cycle', 'Workflow', 'Application development', 'Na
3	Experienced Data	Volvo Financia	4.6	3-8 Yrs	Not disclo	Kolkata, Mumbai, New De	You will work in agile teams through good collaboration with our colleagues in ...	['Automation', 'C++', 'Powertrain', 'Mechatronics', 'Simulation', 'Construction equipment', 'Simulink', 'MATLAB']
4	Data Scientist - Ki	Capco	3.7	8-13 Yrs	Not disclo	Kolkata, Mumbai, New De	Requirements and Skills : - 8+ years of experience in data science field Experience in...	['Team management', 'Prototype', 'Database design', 'Agile', 'Data structures', 'MATLAB', 'Analytics', 'Financial services']
5	Data Scientist - Re	Wipro	3.8	5-10 Yrs	Not disclo	Pune, Bengaluru	Shift Required: An undergraduate degree in statistics, mathematics, computer science, e...	['Reinforcement learning', 'Q Learning', 'Deep Learning', 'Data', 'Q', 'Reinforcement']
6	AI/ML Data Scien	Ford	4.4	6-10 Yrs	Not disclo	Kolkata, Mumbai, New De	Bachelors or Masters degree in a related field (e.g., Computer Science, Data Science,...	['Computer science', 'data cleansing', 'Data analysis', 'Machine learning', 'Agile', 'Data processing', 'data visualization', 'Cont
7	Staff Data Scienti	Visa Inc	4.2	0-3 Yrs	Not disclo	Bangalore/Bengaluru	Basic Qualifications: 5+ years of relevant work experience with a Bachelor s Degree or ...	['Unix', 'Automation', 'Linux', 'Data structures', 'Perl', 'Business intelligence', 'Troubleshooting', 'Open source']
8	Senior Data Scien	Fractal Analyti		4 6-10 Yrs	Not disclo	Mumbai, Pune, Gurgaon/	Experience with programming languages such as Python, R, Spark . Strong understanding o...	['RCA', 'Software design', 'Version control', 'Coding', 'Machine learning', 'Data structures', 'Project delivery', 'Monitoring']
9	VP- Data Scientis	Citicorp Financ	3.8	12-16 Yrs	Not disclo	Hybrid - Pune, Bangalore/	Basic QualificationsDegree in computer science, engineering, mathematics, operations r...	['Data Science', 'Natural Language Processing', 'data scientist', 'Machine Learning', 'Deep Learning', 'VP', 'Data', 'Machine']
10	Data Scientist	Tech Mahindra	3.7	5-10 Yrs	Not disclo	Bangalore/Bengaluru	Job description : Data scientist who works in python with advanced statistical modellin...	['Statistical modeling', 'Deployment', 'Python', 'Testing', 'Software testing', 'Data', 'Modeling', 'Statistics']
11	Data Scientist	Narayana Heal	4.2	3-5 Yrs	Not disclo	Hybrid - Bengaluru, Karna	The candidate should be comfortable in Python and should have worked on data science pr	['Deep Learning', 'Python', 'SQL', 'TensorFlow', 'PyTorch', 'Data']
12	Artificial Intellige	Accenture		4 5-8 Yrs	Not disclo	Bangalore/Bengaluru	Skill required: Artificial Intelligence Innovation - Data Science. You are required to ...	['Data Science', 'Artificial Intelligence Innovation', 'deep learning', 'python', 'nlp', 'client relationship', 'creative solutions', '
13	Data Scientist	IBM	4.1	7-9 Yrs	Not disclo	Bangalore/Bengaluru	Required Technical and Professional Expertise . 7+ years of experienceSound knowledge o.	['Data Science', 'Data analytics', 'Python', 'algorithms', 'data modeling', 'Hadoop', 'statistical modeling', 'Analytics']
14	Data Scientist	Tech Mahindra	3.7	5-8 Yrs	17-25 Lacs	Hybrid - Hyderabad/Secu	Data Scientist. PAN India. Notice Period: Immediate to 30 days only. Must haveExperienc...	['Azure', 'NLP', 'Python', 'Docker', 'FAST API', 'Data', 'API', 'Microsoft Azure']
15	Data Scientist	Brace Infotech		4-9 Yrs	20-35 Lacs	Remote	At least 4+ years of relevant experience as a data scientist 2+ years of experience wor...	['Data Science', 'Python', 'Science', 'Data']
16	Data Scientist - A	Genpact	3.9	4-7 Yrs	Not disclo	Bengaluru	We are looking for data scientists responsible for solving business problems like forec...	['Neural Networks', 'Machine Learning', 'Python', 'Regression', 'Clustering', 'Networking', 'Data', 'Assistance']
17	Data Science Adv	Dell Technolog	4.1	5-7 Yrs	Not disclo	Bangalore/Bengaluru	Essential Requirements Hands-on experience building AI / ML models and working directly	['Data Science', 'PyTorc', 'SciPy', 'Artificial Intelligence', 'Plotly', 'Natural Language Processing', 'AI/ML models', 'Scikit-Learn'
18	Data Scientist	Infosys	3.9	2-6 Yrs	Not disclo	Bangalore/Bengaluru	Bachelor of Engineering experience in optimization (linear, mixed integer, constraint ...	['Telecom', 'SAS', 'Billing', 'SQL', 'Simulation', 'SPSS', 'Business intelligence', 'Forecasting']
19	Data Scientist	IBM	4.1	8-12 Yrs	Not disclo	Bangalore/Bengaluru	The candidate should possess a strong background in Python programming, database manag	['CSS', 'JavaScript', 'HTML', 'React', 'open source', 'python', 'predictive modeling', 'web technologies']
20	Data Scientist - Se	Samsung India	4.5	6-8 Yrs	Not disclo	Bangalore/Bengaluru	Bachelors or Masters degree in Computer Science, Statistics, Mathematics, or a related ...	['Data analysis', 'Staffing', 'Analytical', 'Machine learning', 'Business strategy', 'data visualization', 'SQL', 'Data extraction']
21	Data Scientist	PwC	3.6	4-9 Yrs	12-22 Lacs	Hybrid - Bangalore/ Beng	Data Science Manager:Data Science Manager - 7 - 9 Years of ExperienceJob Description:O...	['Data Science', 'Machine Learning', 'Python', 'Science', 'Data', 'Machine']
22	Data Scientist/ Se	Fractal Analyti		4 4-9 Yrs	Not disclo	Mumbai, Pune, Chennai,	Demonstrate excellent communication skills and the ability to work effectively in a tea...	['Computer science', 'GIT', 'GCP', 'Analytical', 'Software development life cycle', 'Workflow', 'Application development', 'Na
23	Experienced Data	Volvo Financia	4.6	3-8 Yrs	Not disclo	Kolkata, Mumbai, New De	You will work in agile teams through good collaboration with our colleagues in ...	['Automation', 'C++', 'Powertrain', 'Mechatronics', 'Simulation', 'Construction equipment', 'Simulink', 'MATLAB']
24	Data Scientist - Ki	Capco	3.7	8-13 Yrs	Not disclo	Kolkata, Mumbai, New De	Requirements and Skills : - 8+ years of experience in data science field Experience in...	['Team management', 'Prototype', 'Database design', 'Agile', 'Data structures', 'MATLAB', 'Analytics', 'Financial services']
25	Data Scientist - Re	Wipro	3.8	5-10 Yrs	Not disclo	Pune, Bengaluru	Shift Required: An undergraduate degree in statistics, mathematics, computer science, e...	['Reinforcement learning', 'Q Learning', 'Deep Learning', 'Data', 'Q', 'Reinforcement']
26	AI/ML Data Scien	Ford	4.4	6-10 Yrs	Not disclo	Kolkata, Mumbai, New De	Bachelors or Masters degree in a related field (e.g., Computer Science, Data Science,...	['Computer science', 'data cleansing', 'Data analysis', 'Machine learning', 'Agile', 'Data processing', 'data visualization', 'Cont
27	Staff Data Scienti	Visa Inc	4.2	0-3 Yrs	Not disclo	Bangalore/Bengaluru	Basic Qualifications: 5+ years of relevant work experience with a Bachelor s Degree or ...	['Unix', 'Automation', 'Linux', 'Data structures', 'Perl', 'Business intelligence', 'Troubleshooting', 'Open source']
28	Senior Data Scien	Fractal Analyti		4 6-10 Yrs	Not disclo	Mumbai, Pune, Gurgaon/	Experience with programming languages such as Python, R, Spark . Strong understanding o...	['RCA', 'Software design', 'Version control', 'Coding', 'Machine learning', 'Data structures', 'Project delivery', 'Monitoring']
29	VP- Data Scientis	Citicorp Financ	3.8	12-16 Yrs	Not disclo	Hybrid - Pune, Bangalore/	Basic QualificationsDegree in computer science, engineering, mathematics, operations r...	['Data Science', 'Natural Language Processing', 'data scientist', 'Machine Learning', 'Deep Learning', 'VP', 'Data', 'Machine']
30	Data Scientist	Tech Mahindra	3.7	5-10 Yrs	Not disclo	Bangalore/Bengaluru	Job description : Data scientist who works in python with advanced statistical modellin...	['Statistical modeling', 'Deployment', 'Python', 'Testing', 'Software testing', 'Data', 'Modeling', 'Statistics']
31	Data Scientist	Narayana Heal	4.2	3-5 Yrs	Not disclo	Hybrid - Bengaluru, Karna	The candidate should be comfortable in Python and should have worked on data science pr	['Deep Learning', 'Python', 'SQL', 'TensorFlow', 'PyTorch', 'Data']
32	Artificial Intellige	Accenture		4 5-8 Yrs	Not disclo	Bangalore/Bengaluru	Skill required: Artificial Intelligence Innovation - Data Science. You are required to ...	['Data Science', 'Artificial Intelligence Innovation', 'deep learning', 'python', 'nlp', 'client relationship', 'creative solutions', '
33	Data Scientist	IBM	4.1	7-9 Yrs	Not disclo	Bangalore/Bengaluru	Required Technical and Professional Expertise . 7+ years of experienceSound knowledge o.	['Data Science', 'Data analytics', 'Python', 'algorithms', 'data modeling', 'Hadoop', 'statistical modeling', 'Analytics']
34	Data Scientist	Tech Mahindra	3.7	5-8 Yrs	17-25 Lacs	Hybrid - Hyderabad/Secu	Data Scientist. PAN India. Notice Period: Immediate to 30 days only. Must haveExperienc...	['Azure', 'NLP', 'Python', 'Docker', 'FAST API', 'Data', 'API', 'Microsoft Azure']
35	Data Scientist	Brace Infotech		4-9 Yrs	20-35 Lacs	Remote	At least 4+ years of relevant experience as a data scientist 2+ years of experience wor...	['Data Science', 'Python', 'Science', 'Data']
36	Data Scientist - A	Genpact	3.9	4-7 Yrs	Not disclo	Bengaluru	We are looking for data scientists responsible for solving business problems like forec...	['Neural Networks', 'Machine Learning', 'Python', 'Regression', 'Clustering', 'Networking', 'Data', 'Assistance']

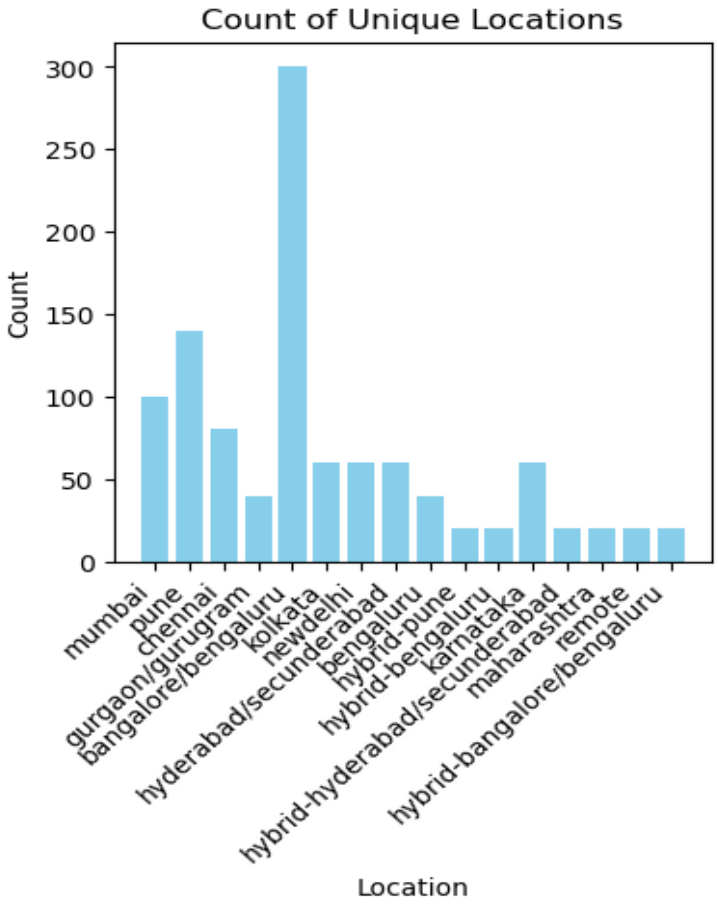
Data Sample After Cleaning Process

After initially collecting 600 samples from the Naukri job portal, a thorough data cleaning process was carried out to create a high-quality dataset. This involved removing duplicates and extraneous items, resulting in a refined dataset of 400 samples. The multifaceted cleaning procedure addressed issues like missing values and discrepancies, ensuring dataset integrity. This rigorous approach aimed to enhance reliability and accuracy for subsequent analyses, contributing to robust findings. The refined dataset, now manageable and well-prepared, underwent further exploration through data visualization with a representative sample of 400 data points.

Market Data Visualization:

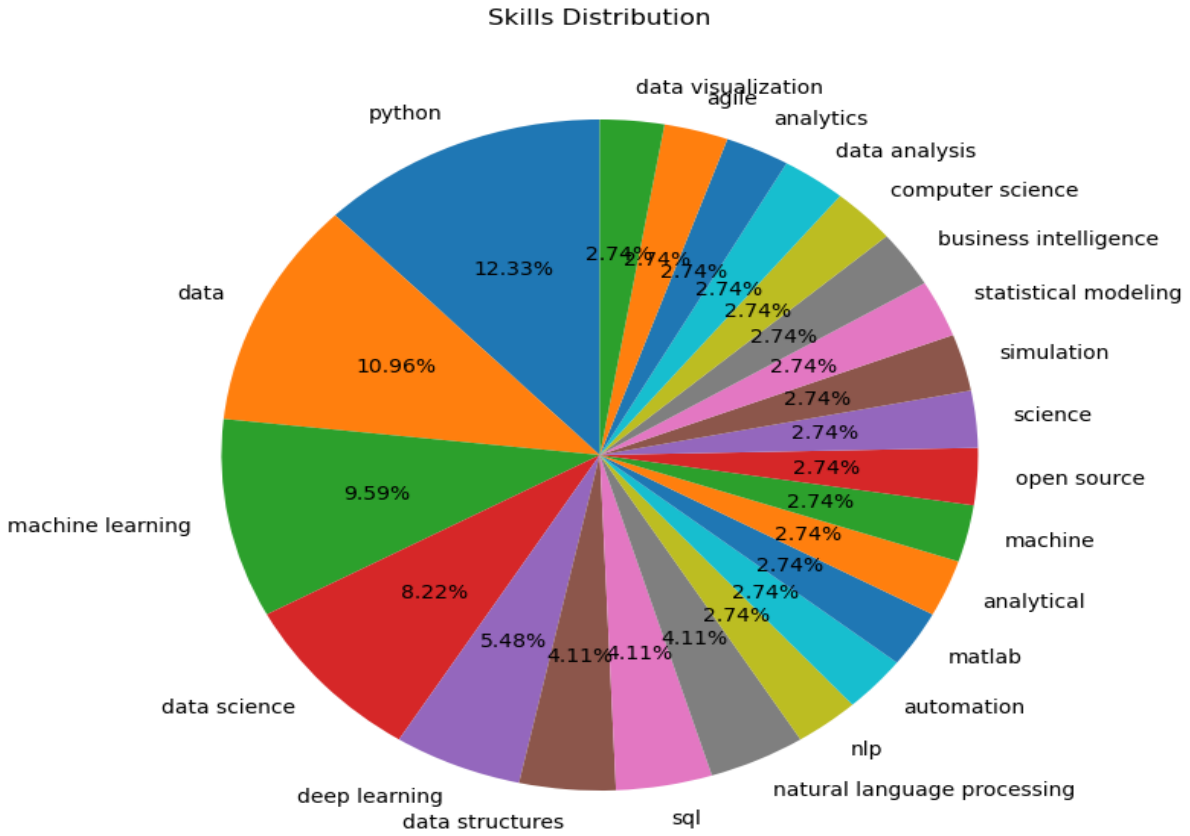
Location Analysis:

The bar graph illustrates the geographical distribution of open job positions in India, with Bengaluru standing out as the predominant hub for Data Scientist roles. Following closely are Pune, Mumbai, and Chennai, showcasing significant opportunities in these cities as well. Interestingly, the categories of "Hybrid" and "Remote" have gained prominence, underscoring the profound impact of the post-COVID work landscape on job preferences and the increasing acceptance of flexible work arrangements.



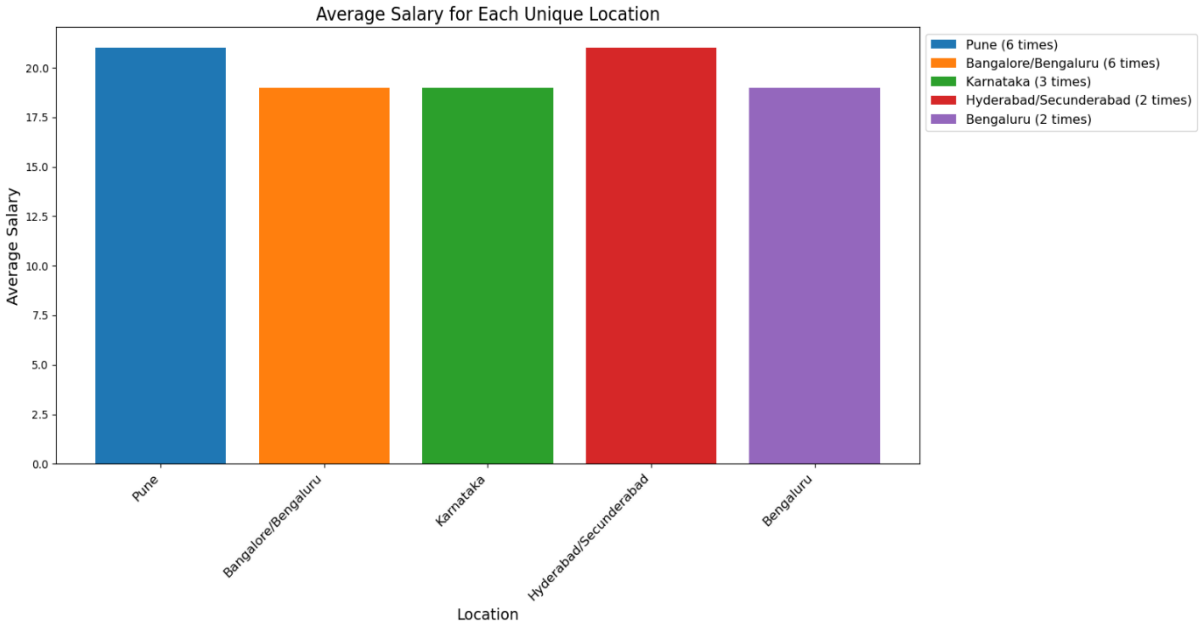
Skill-set Analysis:

Python's essential importance in Data Science was highlighted graphically, shaping the field. Python proficiency is required, as are skills in machine learning, deep learning, and complex data structures. Expertise in Database Management Systems, particularly SQL, was emphasized as essential for effective data handling. Essential abilities such as Business Intelligence, statistical modelling, and analytical expertise were also highlighted in the dynamic arena of machine learning and artificial intelligence applications in Data Science.

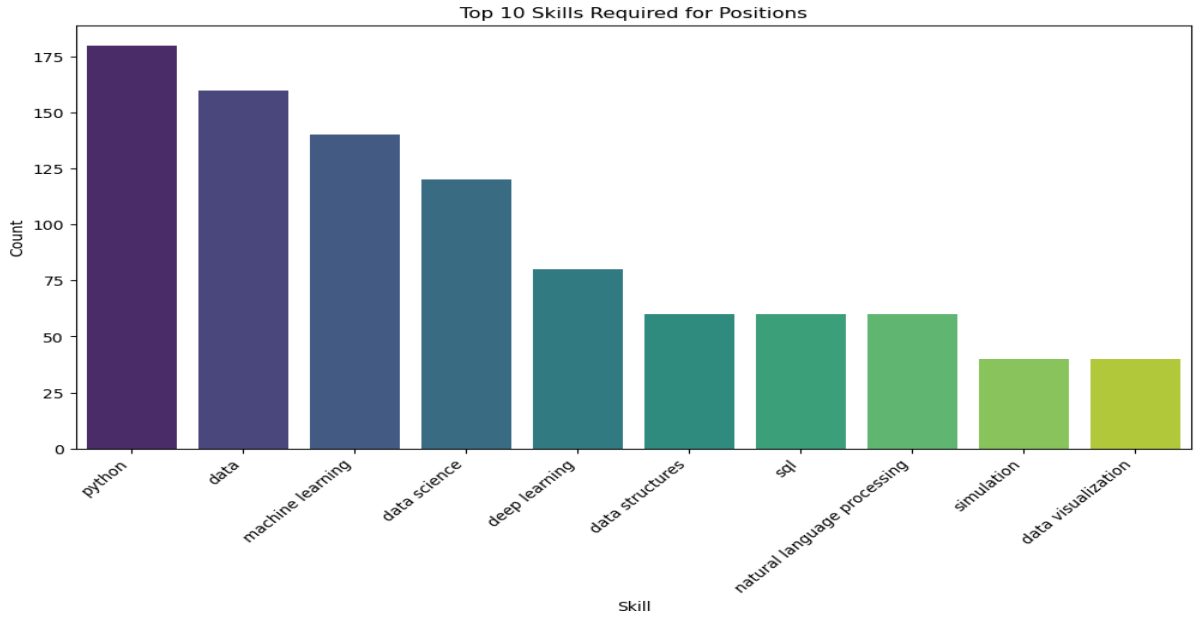


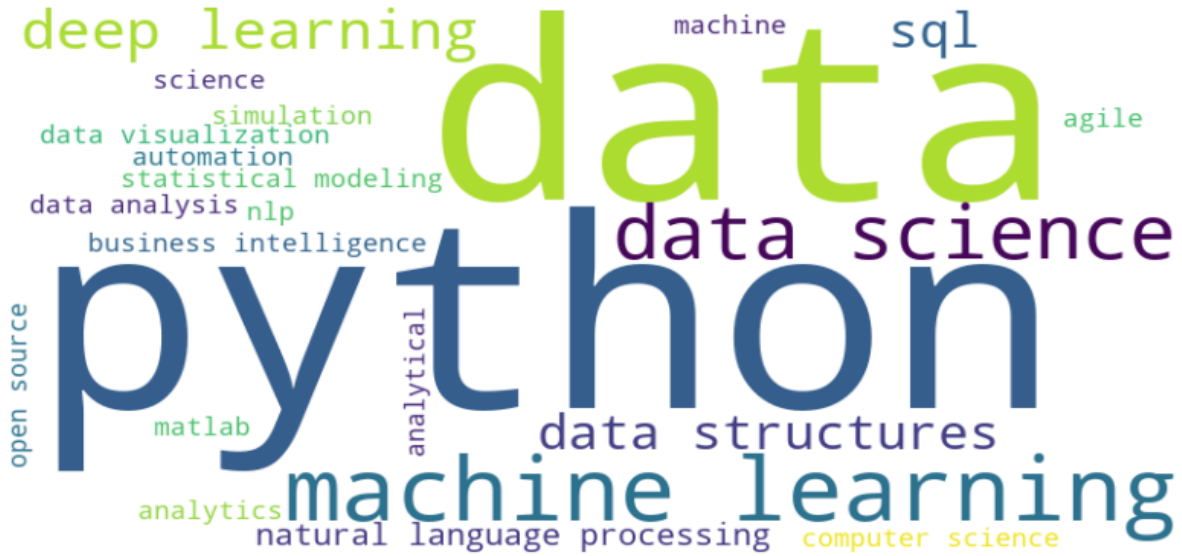
Salary Analysis:

While a large chunk of the dataset lacked wage information tagged as "Not Disclosed," the legitimate values ranged from 1 LPA to 25 LPA. However, due to the small number of items, it is difficult to draw reliable salary-related conclusions from this dataset.



Based on the dataset, additional visual representation graphs were developed, providing additional insights and viewpoints for an extensive understanding of the data.





My Ideal Job:

Title - AI/ML Data Scientist

Skills - 'Computer science', 'data cleansing', 'Data analysis', 'Machine learning', 'Agile', 'Data processing', 'data visualization', 'Continuous improvement'

Aligned with prevailing market trends, this ideal job underscores the importance of a robust foundation in machine learning, diverse technical skills, and adept data visualization. Effective communication and business acumen are integral, reflecting the comprehensive responsibilities of a data scientist in today's dynamic job landscape.