PART A – JOB MARKET ANALYSIS

The aim of this project is to do an in-depth analysis of job market data related to positions in the 'Data Scientist' domain within a particular regional area. This process includes a systematic collection of job posts from several recruitment platforms, an analysis of prominent job categories, visualization of significant data features, and determining the critical skills required for a specific job function.

Data Collection Methodology:

The study analyses job listing data gathered from a job portal. The dataset offers useful insights on the breakdown of job titles, abilities, and other pertinent details. The research seeks to identify patterns and trends in the job market. The dataset comprises a total of 320 job listings. This sample size was collected from naukri.com during the period Nov'23. The data includes various details such as job titles, companies, ratings, salary information, and skills required.

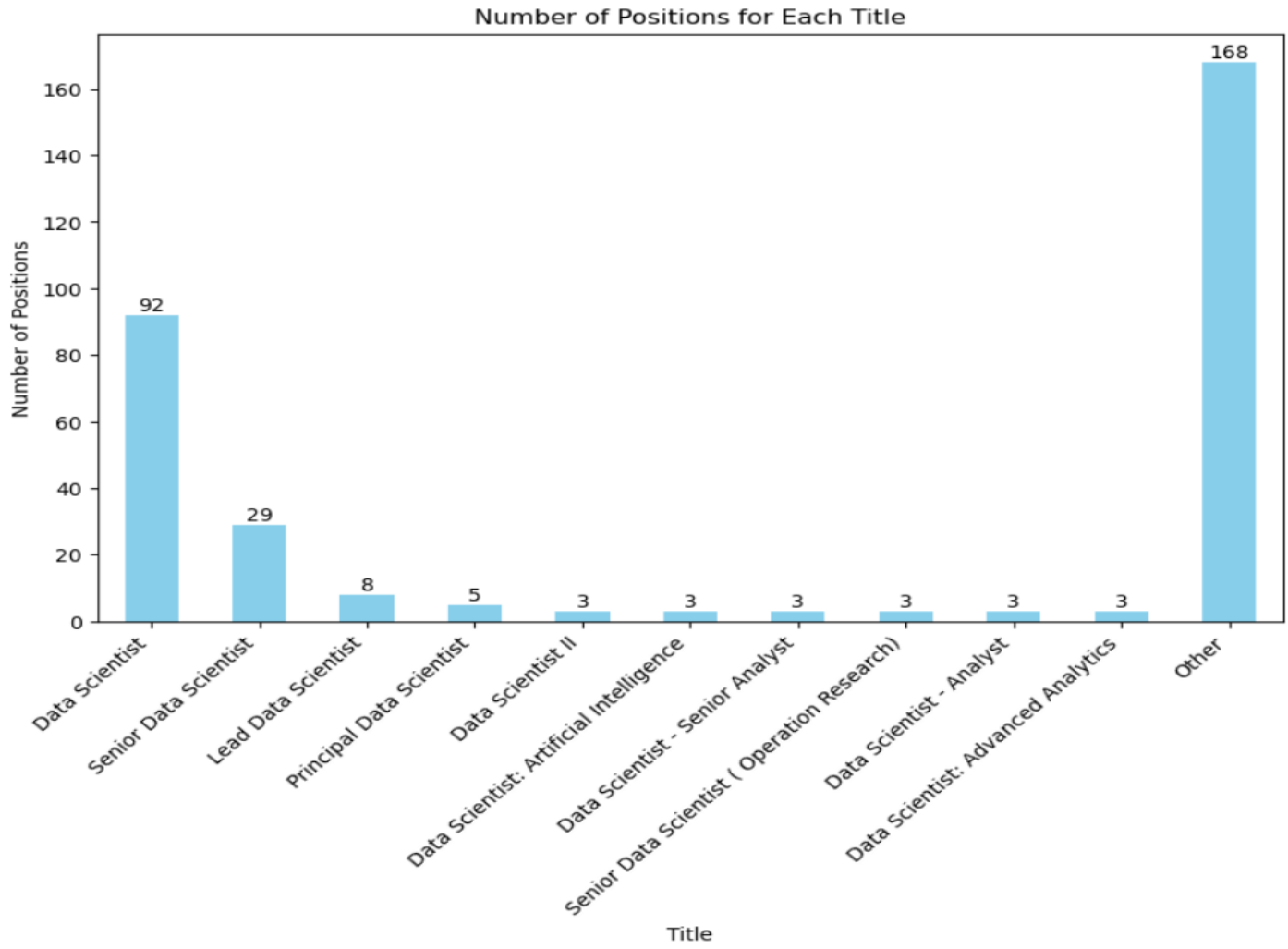
Web scraping from naukri.com was used to collect data, especially job listings relating to data scientist roles. For scraping, the Python programming language and libraries such as Beautiful Soup and Selenium were used. In order to prevent being restricted, proper HTTP headers that replicated the web browser were used. The scraping program was identified as a valid browser using the user-agent string. HTTP GET requests are made to the provided URLs in order to receive the HTML content of the job listing pages. Beautiful soup was used to parse the HTML text and extract useful information.

Identified HTML elements containing job titles, companies, ratings, salary information, skills, and so on, and extracted data from those elements using Beautiful Soup's search and find all methods. The retrieved data was organized into a Pandas Data Frame for easy manipulation and analysis. Based on the retrieved data, suitable column names were assigned to the Data Frame. Checked and handled missing data correctly. The data frame was then exported to a CSV file for future use.

Ensured that ethical web scraping techniques were followed. I followed the robots.txt regulations and avoided sending too many queries in a short period of time. The website employs dynamic content loading, and that issue was addressed using Selenium.

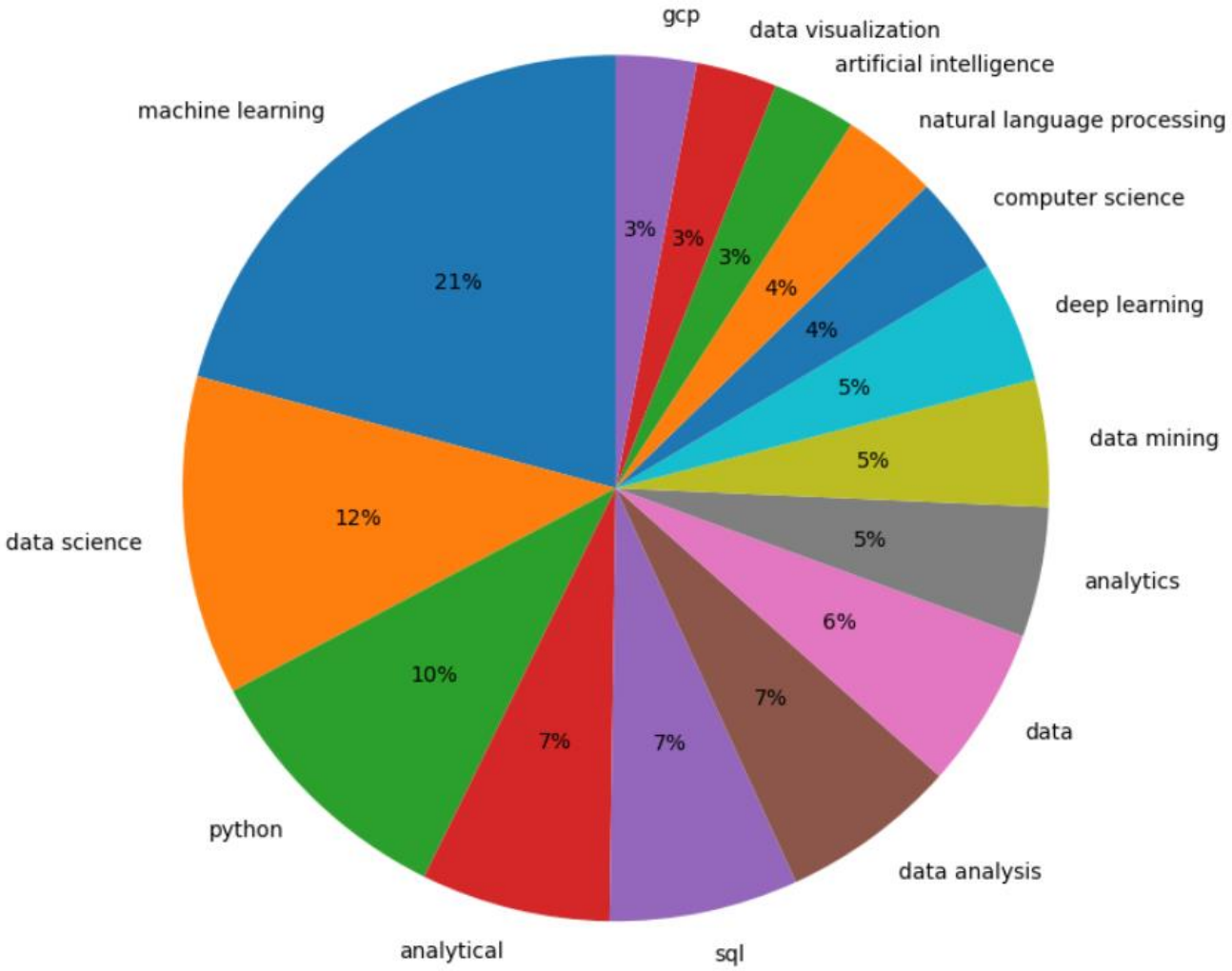
Market Data Visualization

- Titles of positions available



- Multivariate scatter plot for Location and Salary

- Key skills required



Ideal Job:

- Job Title: The conclusion of this rigorous research resulted in the identification of my desired job duties within the enormous world of data and data science after completing an in-depth and thorough analysis. The following are the important talents and abilities necessary for this role:
  - Data Analysis: Proficiency in extracting meaningful insights from large datasets. Ability to clean, pre-process, and manipulate data for analysis.
  - Programming Languages: Strong command of programming languages such as Python, and SQL. Capability to develop and implement algorithms and models.
  - Machine Learning: Knowledge of machine learning algorithms and frameworks. Practical experience in developing and deploying machine learning models.
  - Data Visualization: Skill in creating visualizations to effectively communicate findings. Proficiency in tools like Matplotlib, Seaborn, or Tableau.

Finally, the identification of my ideal job within the vast subject of data science reflects the result of extensive research. The essential skills and abilities required for this role include expertise in data analysis, programming language proficiency, a solid understanding of statistical methods, knowledge of machine learning, adept data visualization skills, and familiarity with database management and big data technologies.

A solid foundation in domain-specific knowledge, problem-solving talents, excellent communication skills, and a dedication to ongoing learning are also required for success in this profession. The combination of these abilities enables me to navigate and contribute significantly to the dynamic and ever-changing field of data science. As I begin this career path, I am well-prepared to solve difficult data issues, extract useful insights, and drive informed decision-making in the field of data and analytics.