Web Scraping Data Science Job Listings

By Rohit Paul

Analysis Steps

Define Objective

Web Scraping

Data Cleaning

Exploratory Data Analysis (EDA)

Identifying Trends and Patterns

Conclusion

Objective

- Develop an intelligent tool that streamlines data science job searches by utilizing web scraping on the Jobs website. Through the extraction of key details and the presentation of insights via visualizations, the tool aims to assist individuals in navigating the data science job market. It also keeps professionals, job seekers, and recruiters well-informed about industry trends.

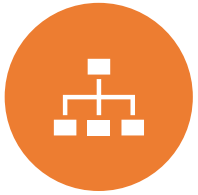
Web Scraping

Started by scraping job info from TimesJobs website.

Used Python code with BeautifulSoup for details.

Extracted job title, company, skills, time, location, and salary.

Kept refining the process for multiple pages.



Managed different data structures on the website.

Data Cleaning

Applied the `strip()` and `replace()` methods systematically using the Pandas library for data organization. Ensured uniformity in data representation by addressing spaces, line breaks, and special characters, enhancing overall readability.

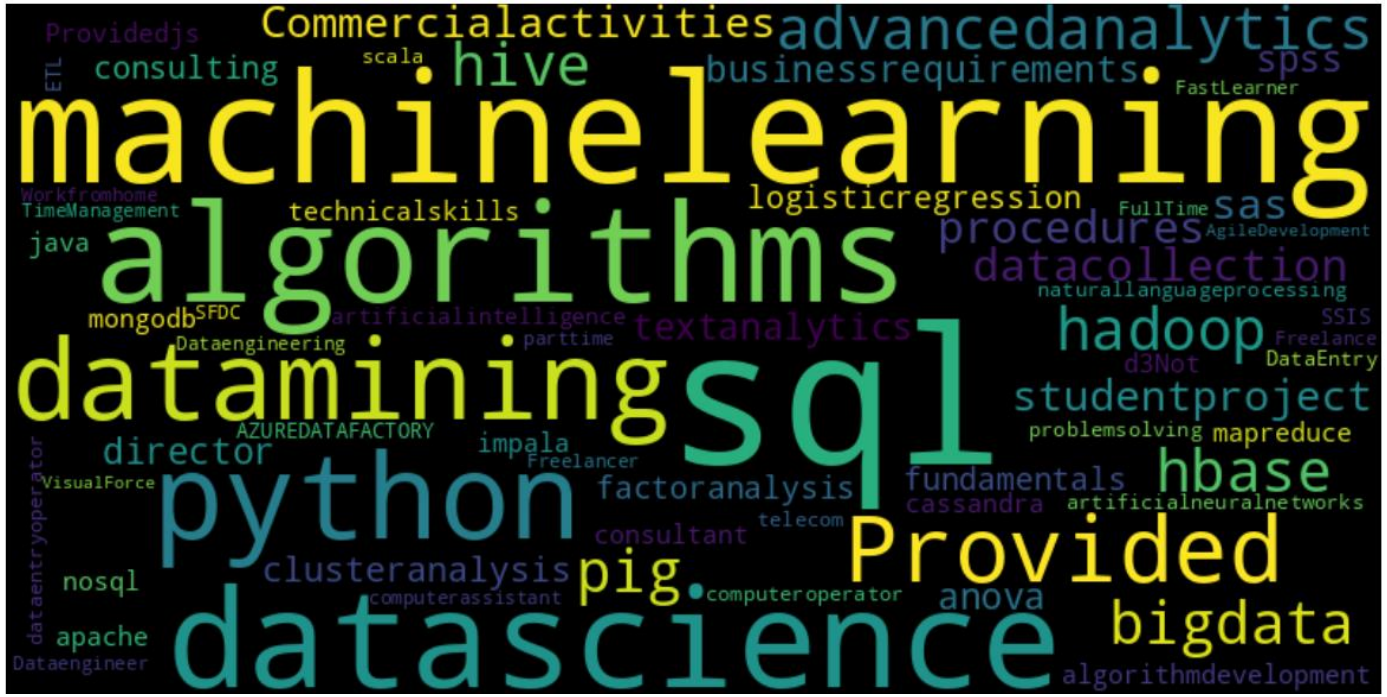
Developed custom functions to ensure a clean extraction of salary information, handling variations like 'Lacs,' and refining experience data for consistency.

Exploratory Data Analysis (EDA)

WordCloud for In-Demand Skills

- Objective: Visualize the most wanted skills.
- Findings: Python, SQL, Machine Learning, and Data Mining were among the most sought-after skills.

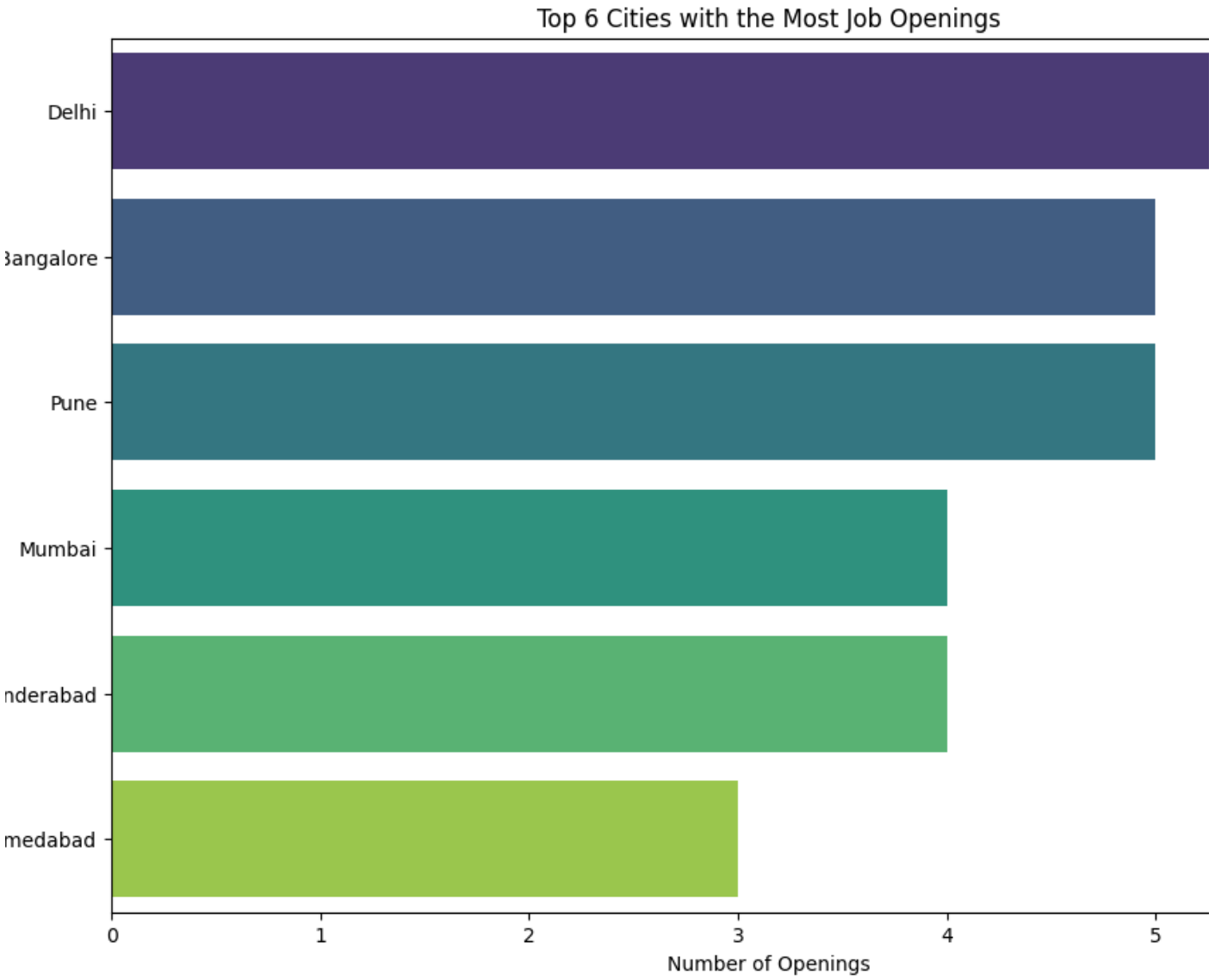
In Demand Skills



Exploratory Data Analysis (EDA)

Top Cities with Most Job Openings

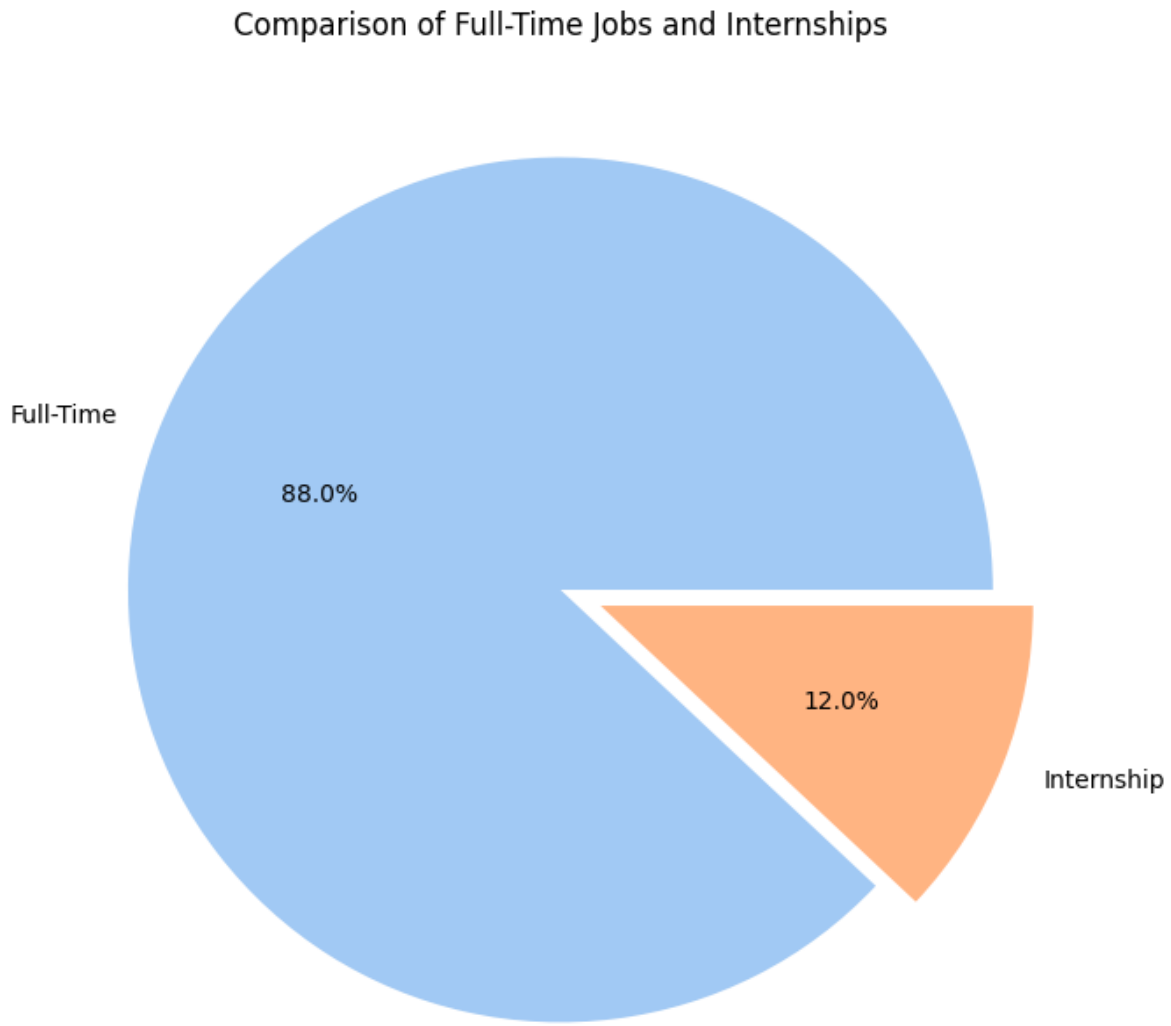
- Objective: Visualize job distribution across cities.
- Findings: Delhi had the highest number of job openings, followed by Bengaluru/ Bangalore, Pune, Mumbai, Hyderabad/Secunderabad and Ahmedabad.



Exploratory Data Analysis (EDA)

Comparison of Full-Time Jobs and Internships

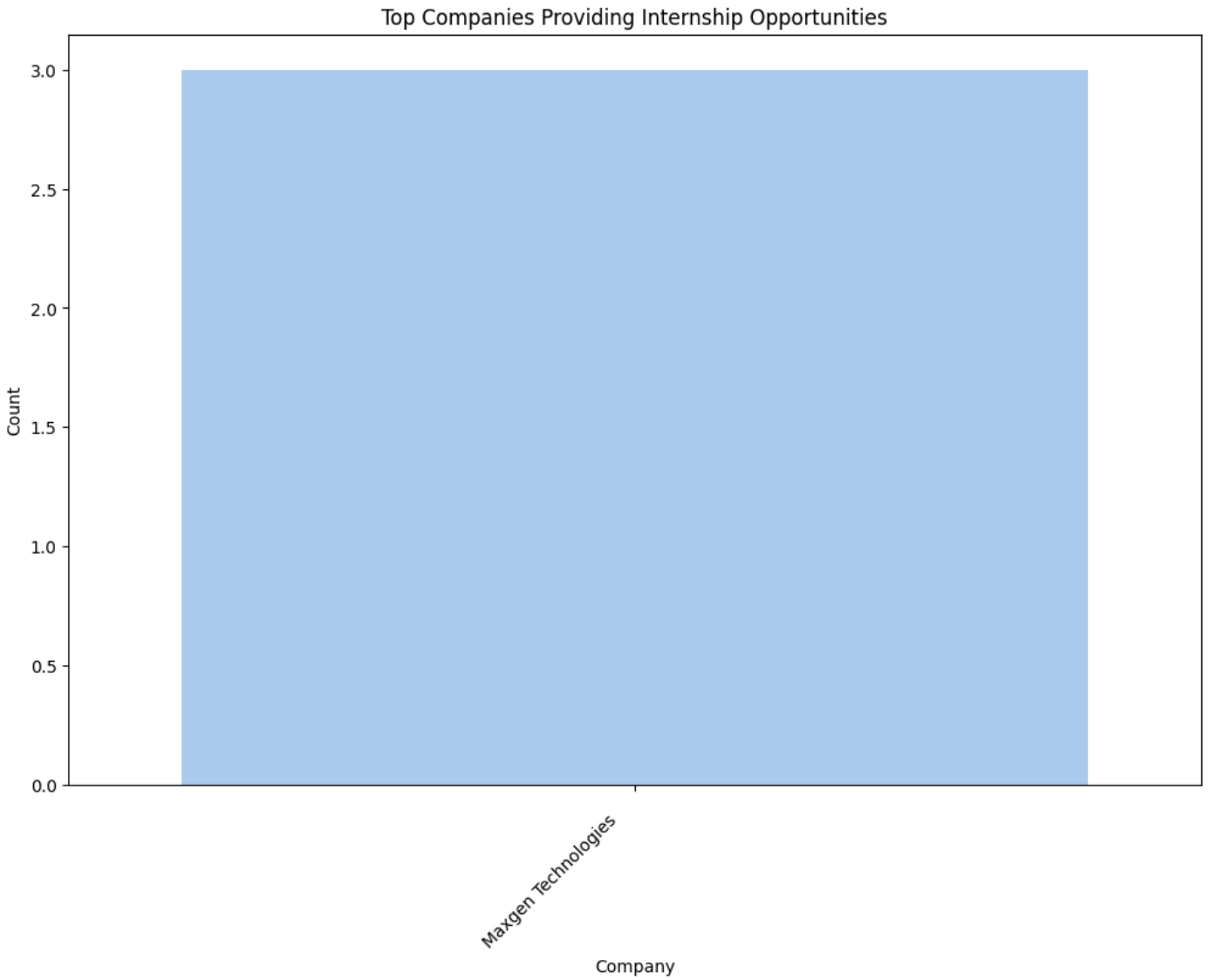
- Objective: Compare the number of full-time jobs with internships.
- Findings: Analyzing the job market revealed that around 88.0% of the opportunities are full-time positions, indicating a predominant demand for full-time roles in the data science field. The remaining 12.0% consists of internship opportunities.



Exploratory Data Analysis (EDA)

Top Companies Providing Internship Opportunities

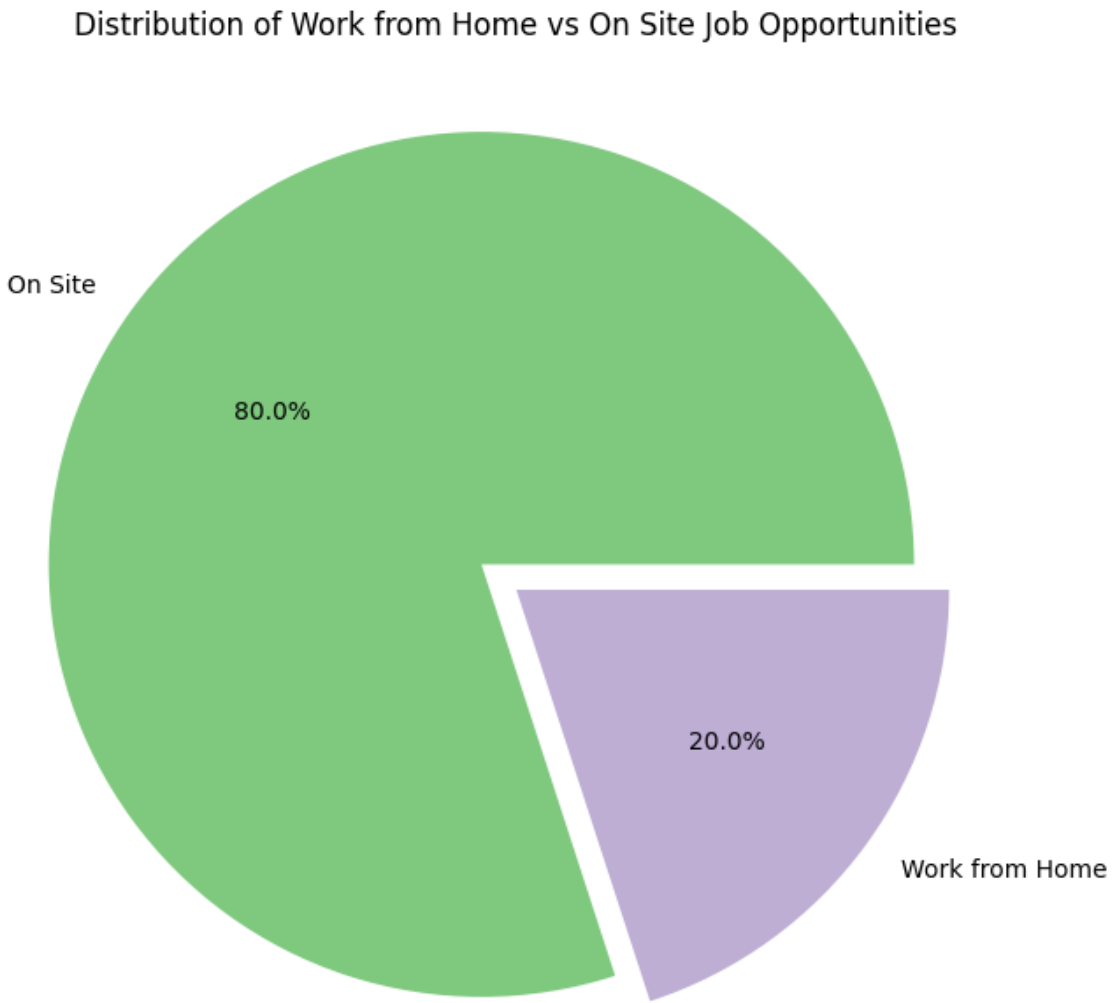
- Objective: Visualize the count of internship opportunities by company.
- Findings: Maxgen Technologies emerged as the top company providing internship opportunities.



Exploratory Data Analysis (EDA)

Comparison of Work from Home vs. On-Site Opportunities

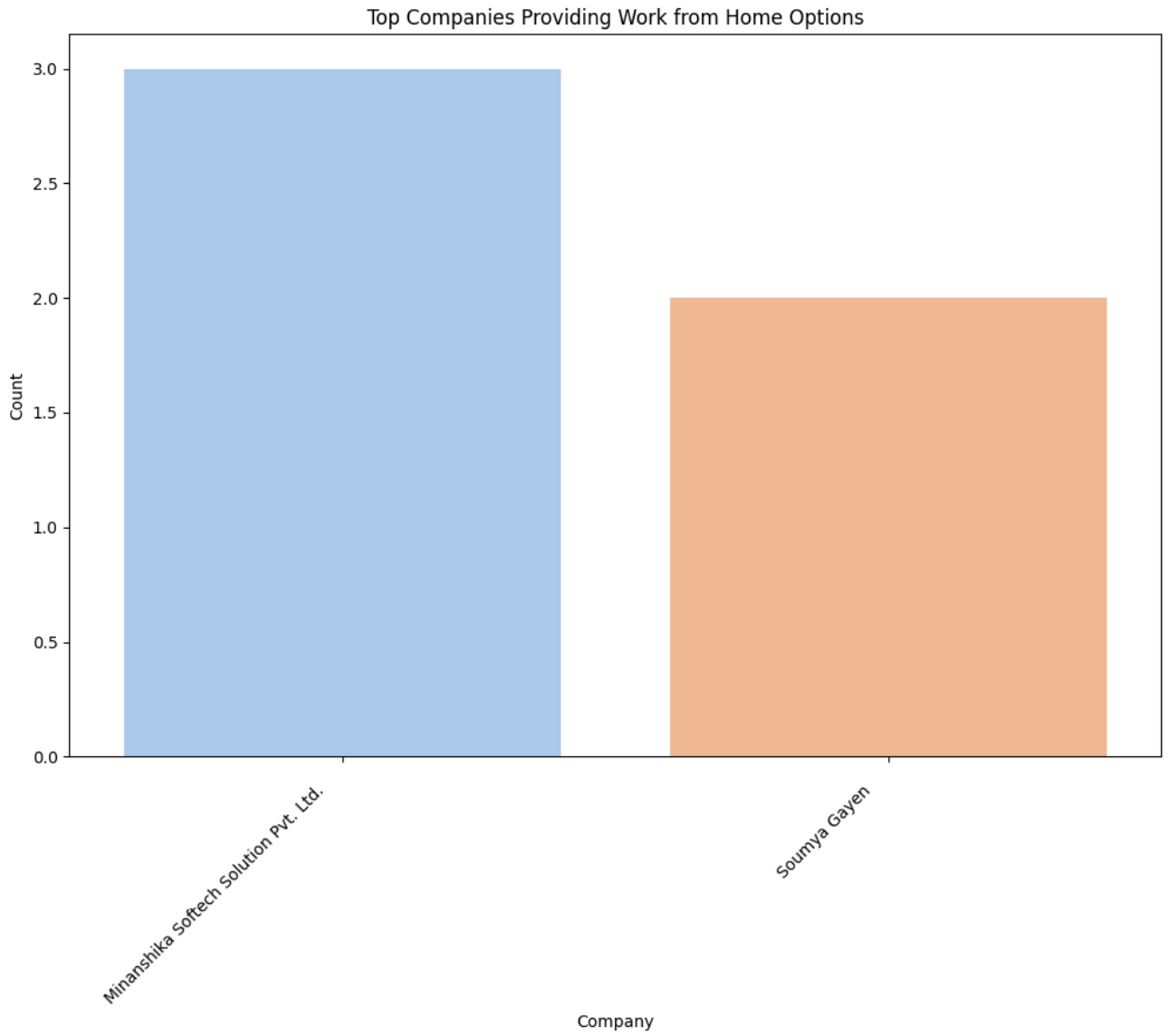
- Objective: Compare the distribution of work-from-home and on-site job opportunities.
- Findings: Analysis of the dataset indicates that 80.0% of jobs require on-site presence, while 20.0% offer work-from-home options, as illustrated in the pie chart.



Exploratory Data Analysis (EDA)

Top Companies Providing Work from Home Options

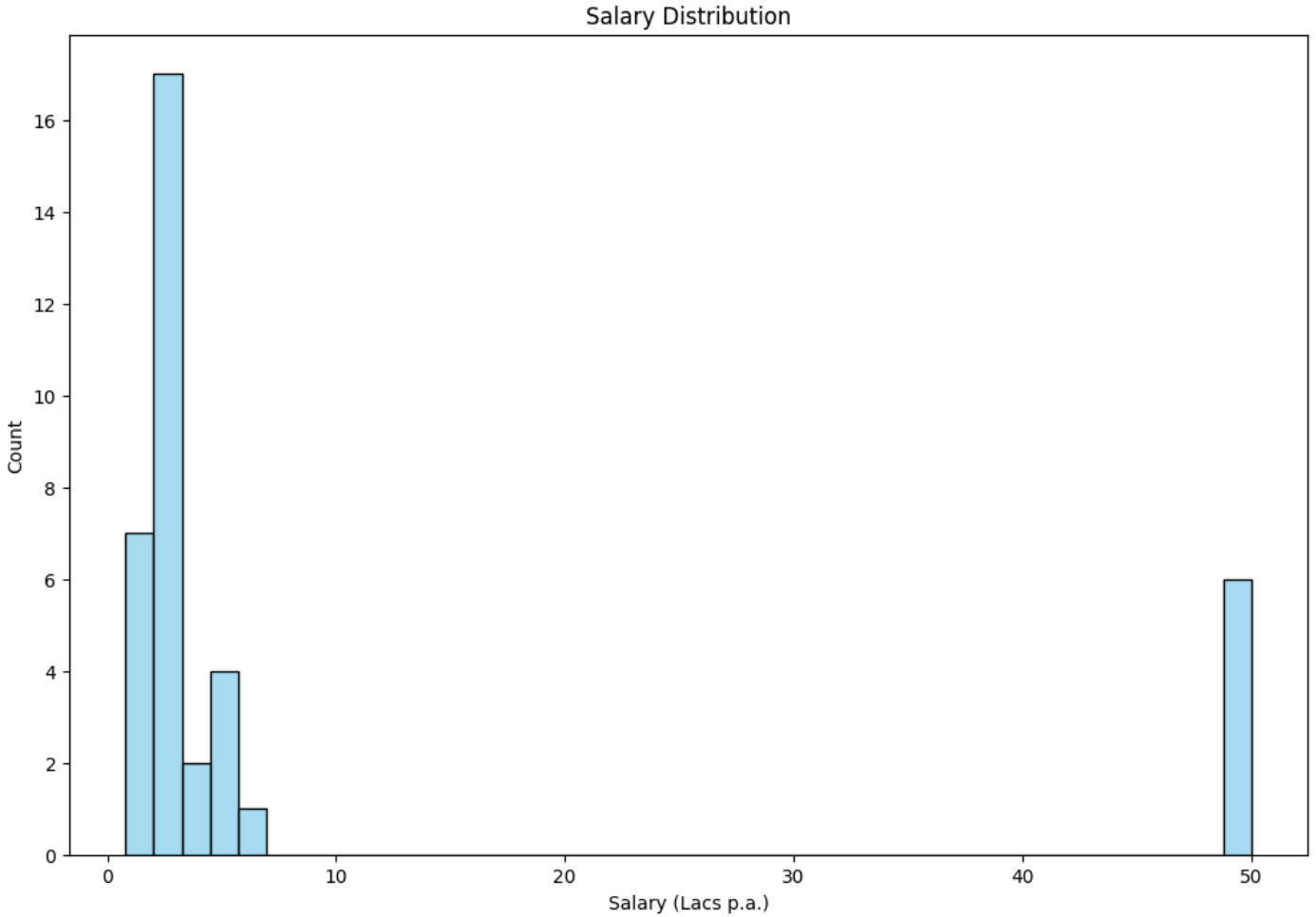
- Objective: Visualize the number of work-from-home opportunities provided by different companies.
- Findings: Minanshika Softech Solution Pvt Ltd leads the list of companies offering work-from-home options, closely followed by Soumya Gayen.



Exploratory Data Analysis (EDA)

Salary Distribution Analysis

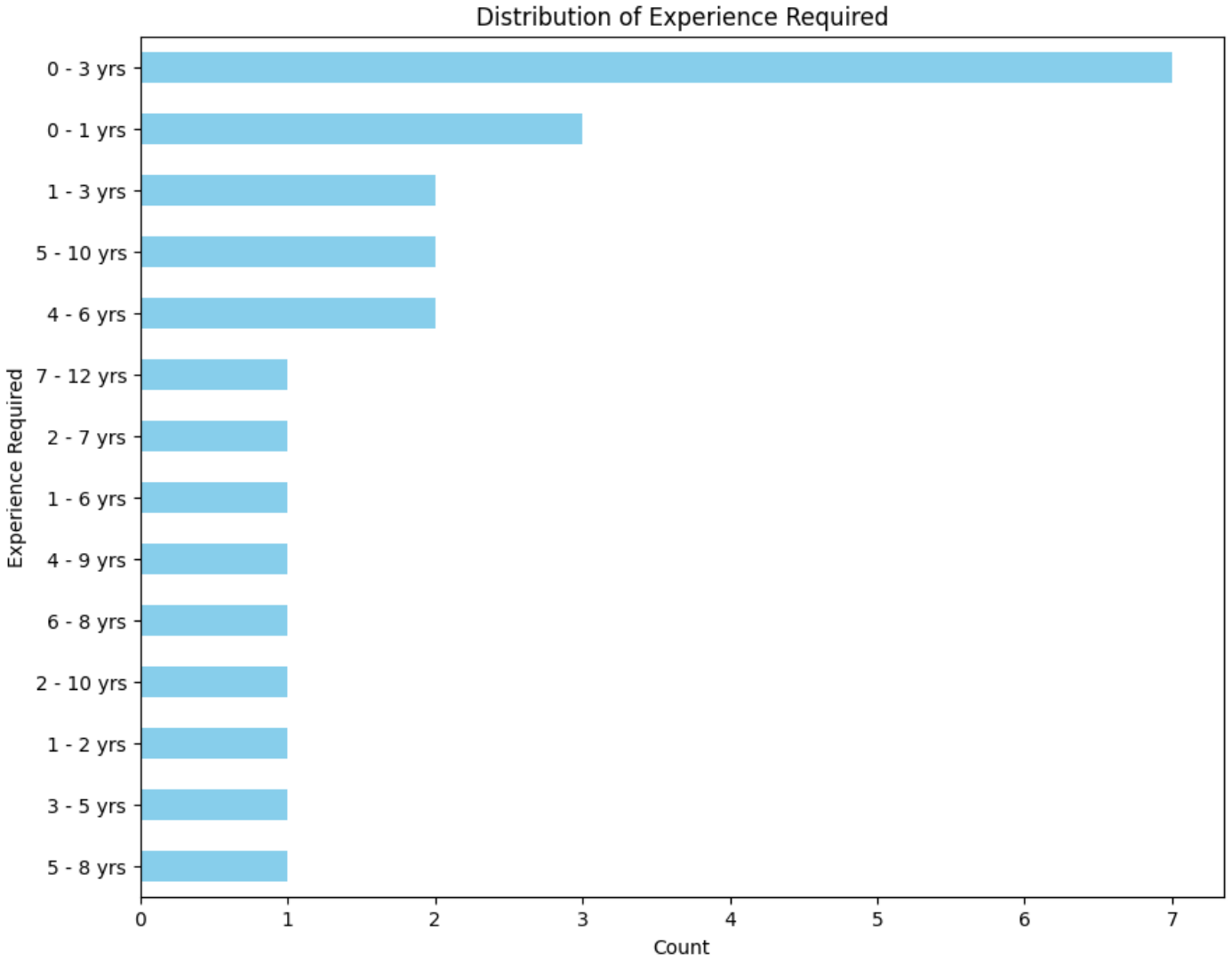
- Objective: Visualize the distribution of salaries.
- Findings: A significant portion of salary ranges centered around 0-10 Lacs per annum, denoting entry-level pay scales, with a more pronounced concentration near 50, reflecting packages designed for experienced professionals.



Exploratory Data Analysis (EDA)

Experience Requirements Analysis

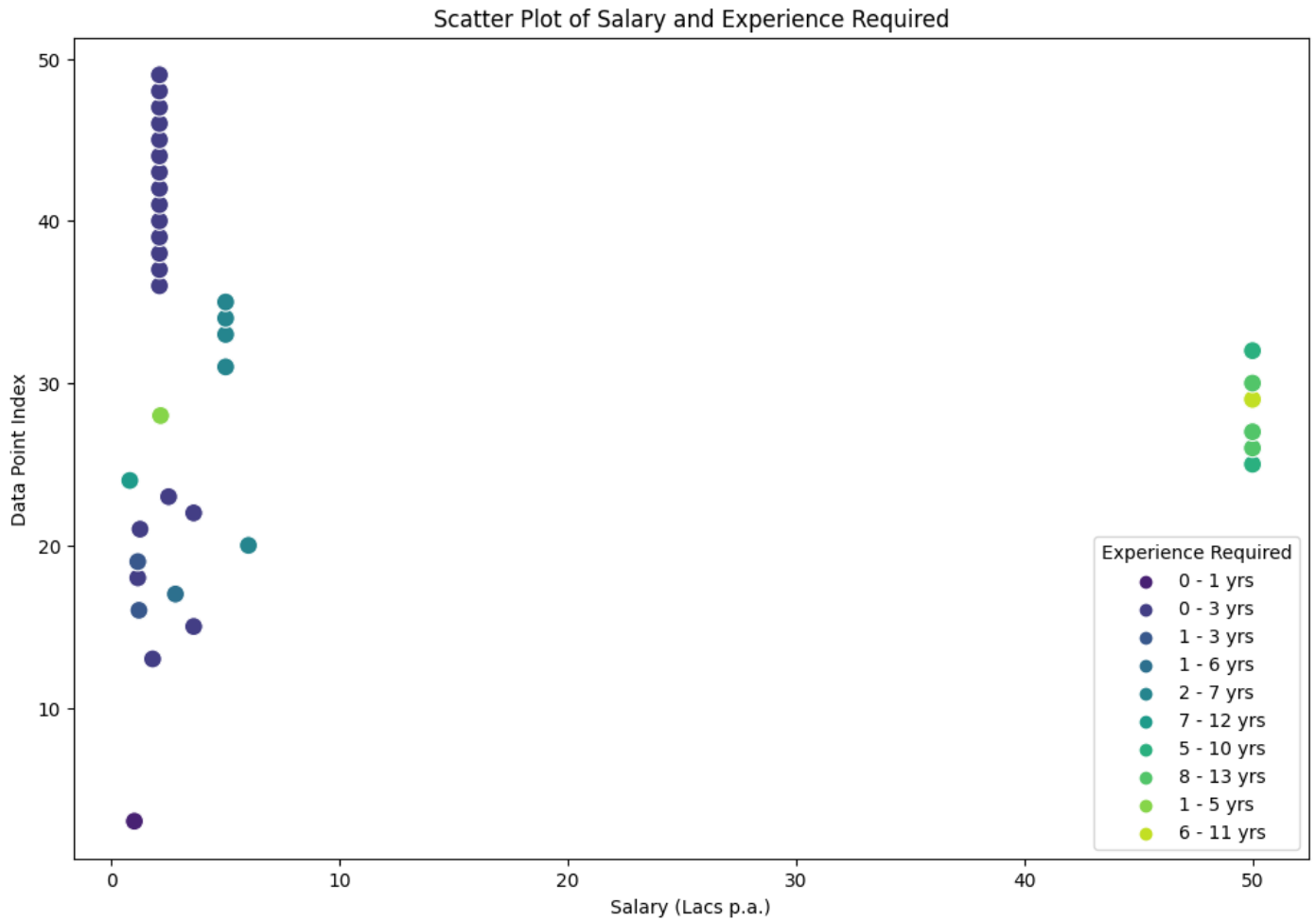
- Objective: Visualize the distribution of required experience levels.
- Findings: In the job market, most jobs are for beginners (0-3 years and 0-1 years experience), followed by roles needing 1-3 years, 5-10 years, 4-6 years experience, and more.



Exploratory Data Analysis (EDA)

Relationship Between Salary and Experience

- Objective: Visualize the relationship between salary and experience.
- Findings: Identified salary clusters corresponding to different experience levels, providing insights into how salary packages (in lacs per annum) vary. A cluster between 0 to 10 indicates predominant entry-level positions, while another cluster near 50 corresponds to higher salary packages for experienced professionals.



Conclusion:

- This project has achieved the development of a valuable tool for exploring the vast landscape of data science job opportunities. Through the extraction of data from the website and the effective use of visualizations, the tool offers crucial insights for professionals, job seekers, and recruiters navigating the dynamic field of data science. It's important to note that these insights are derived from a snapshot of data and may evolve with real-time updates in the ever-changing job market.