Data Analyst Job Market Analysis

Santanu Jha

https://www.linkedin.com/in/santanu-jha-845510292/

Data Analyst Job Strategy

Comprehensive analysis of LinkedIn job postings for Data Analyst roles in India and insights into market trends, skill requirements, and educational qualifications.

OVERVIEW

Project Goal

The primary goal of this project is to simplify the job hunt for Data Analyst positions in India by scraping job listings from LinkedIn, cleaning and preprocessing the data, and performing detailed analysis to provide actionable insights for job seekers.

Key Objectives

- **Data Collection:** Scrape LinkedIn for Data Analyst job postings in India.
- **Data Preprocessing:** Clean and preprocess the collected data.
- **Analysis:** Perform detailed analysis on skills, educational requirements, work modes, job locations, company and industry trends.
- **Hypothesis Testing:** Conduct hypothesis testing to explore relationships between different variables.

DATA WRANGLING

Data Collection

- **Tool Used:** Selenium
- **Source:** LinkedIn
- **Data Saved To:** `AnalystJobs.csv`

Preprocessing and Cleaning

- **Initial Cleaning:** Removed duplicates, separated columns, and standardized formats. Saved to `modified_data.csv`.
- **Advanced Cleaning:** Extracted degree information from job descriptions and saved to `modified_data1.csv`.

Data Dictionary

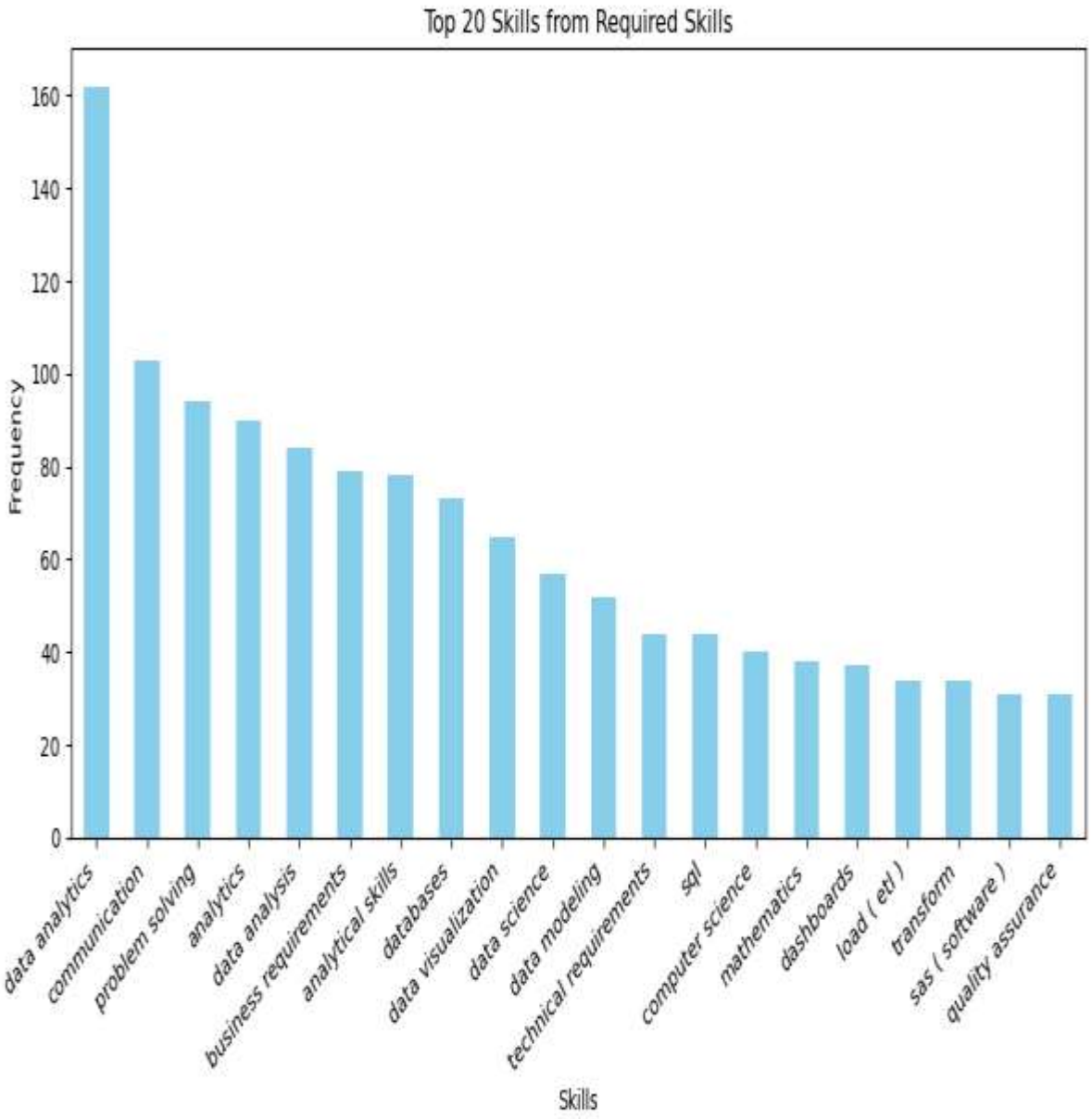
Column Name	Description
Link	URL to the job listing.
Job Title	Title of the job position.
Required Skills	List of skills required for the job, extracted from the job description.
Job Description	Full text of the job description.
Required Experience	Level of experience required for the job (e.g., Entry, Mid, Senior).
Company Name	Name of the company offering the job.
Location	Location of the job, typically formatted as "City, State".
Employee Size	Size of the company in terms of number of employees.
Industry	Industry in which the company operates.
Work Mode	Mode of work (e.g., On-site, Remote, Hybrid).
Actively recruiting	Indicator of whether the company is actively recruiting (Yes/No).
City	City where the job is located, extracted from Location.
State	State where the job is located, extracted from Location.
Cleaned_Job_Description	Preprocessed job description used for text analysis.
bachelor	Binary indicator (0/1) of whether a Bachelor's degree is mentioned.
master	Binary indicator (0/1) of whether a Master's degree is mentioned.

Column Name	Description
phd	Binary indicator (0/1) of whether a PhD degree is mentioned.

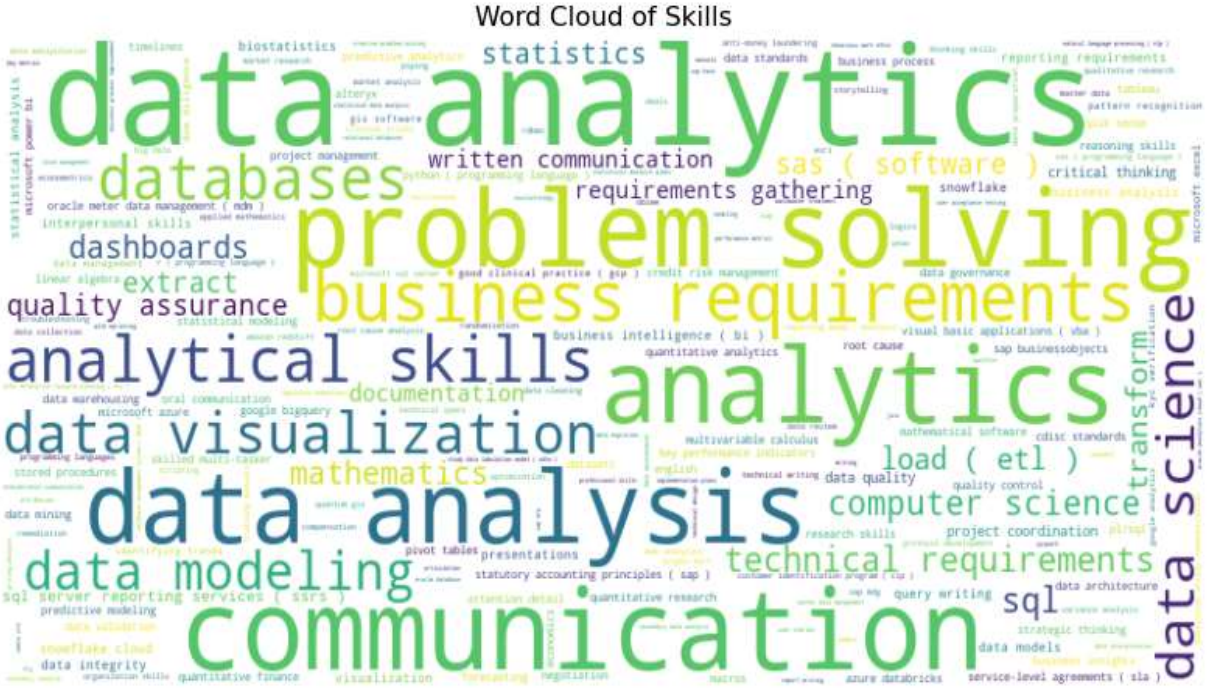
EXPLORATION

Skills Analysis

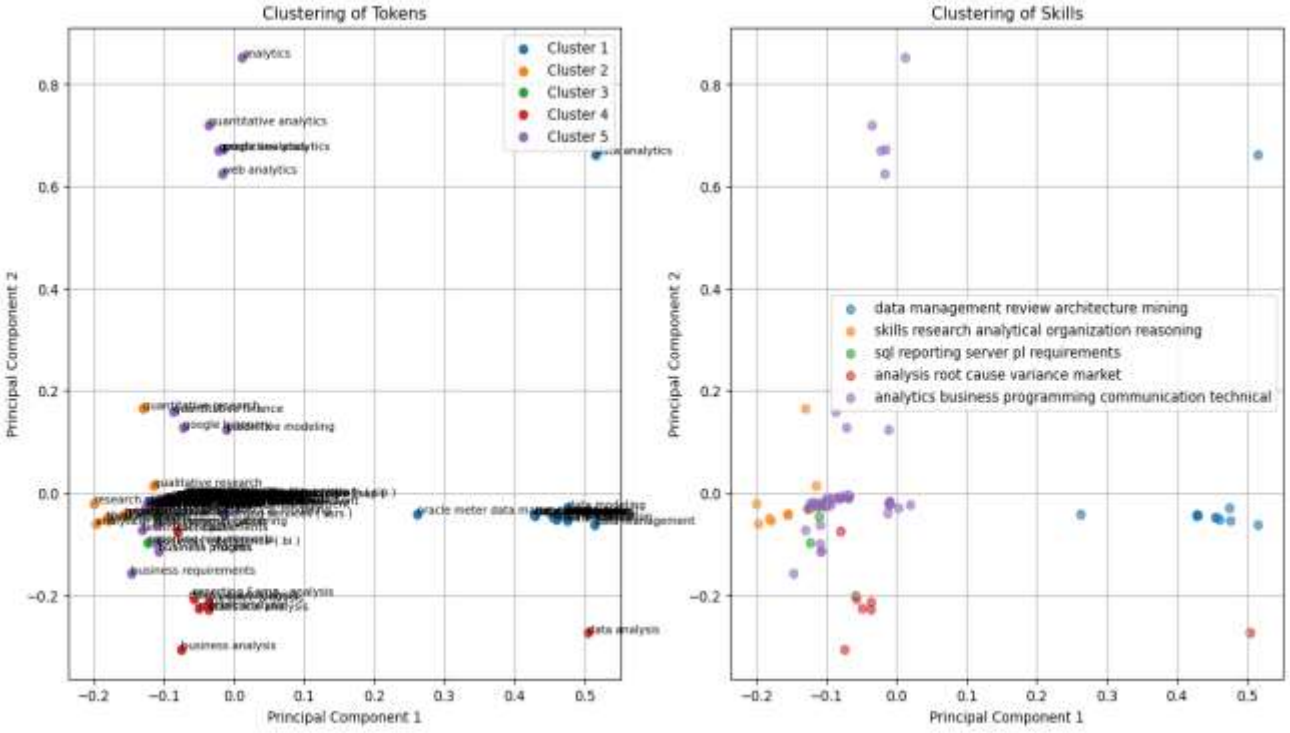
- **Token Frequency:** Identified the most frequently mentioned skills in job listings.



- **Visualization:** Top skills word cloud generated.

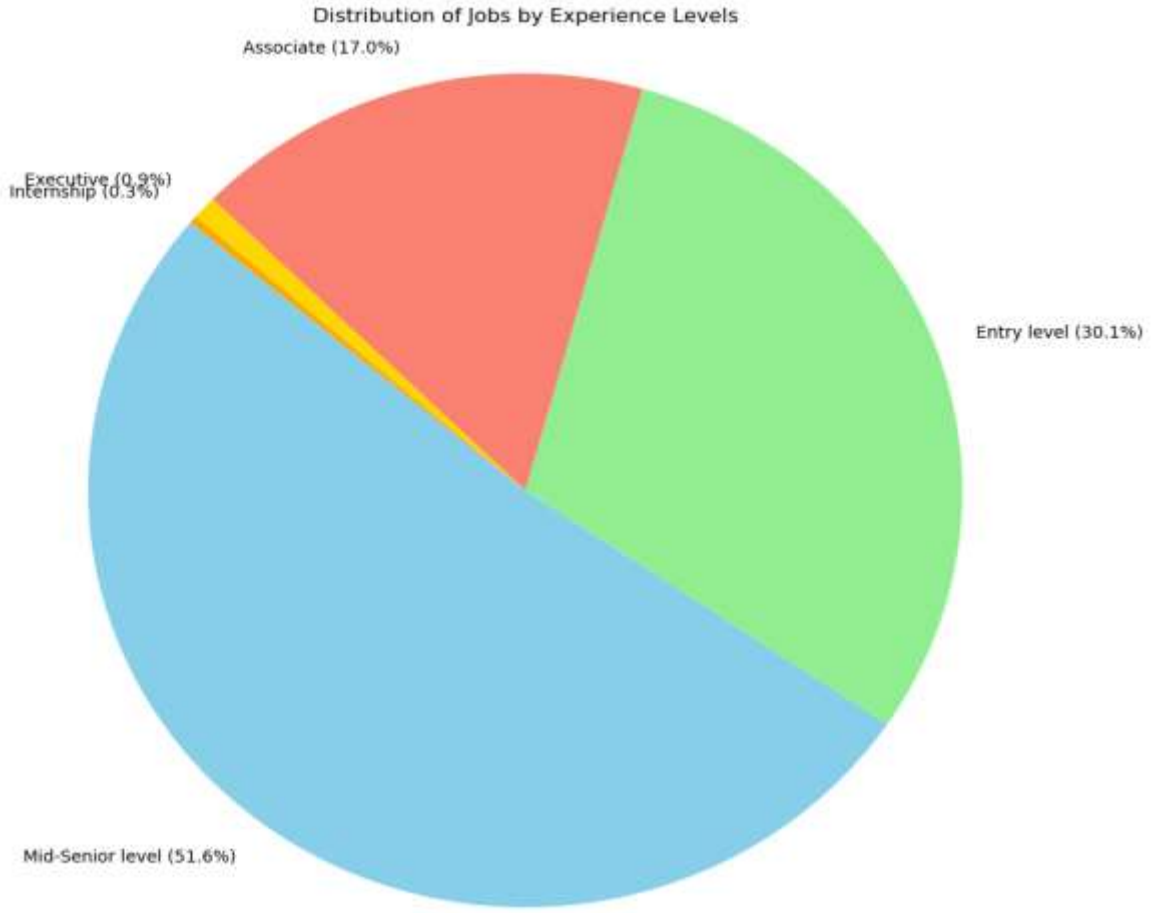


- **Skill Clustering:** Performed using TF-IDF vectorization and K-means clustering. Visualized using PCA.



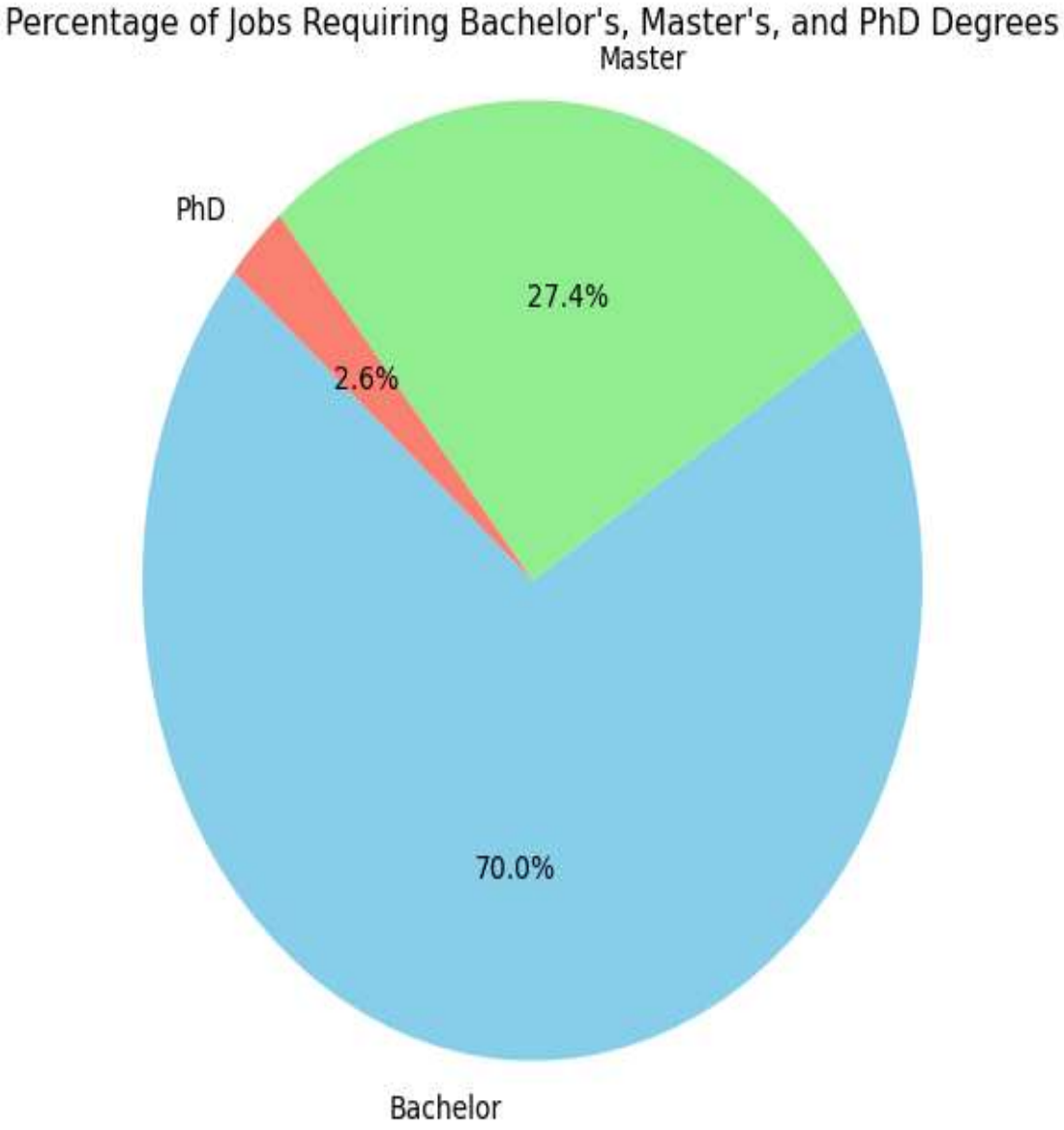
Skills by Experience Level

- Distribution of jobs for different experience levels.

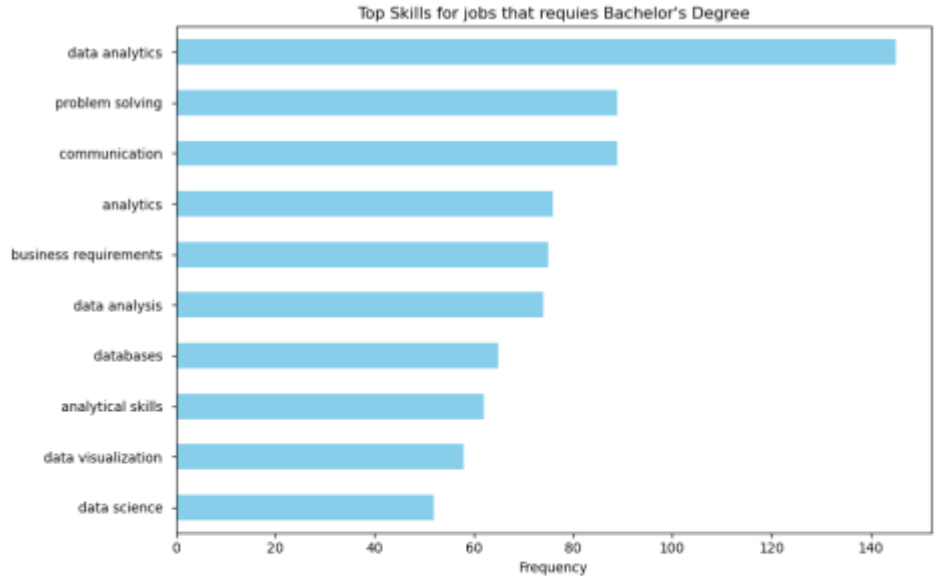


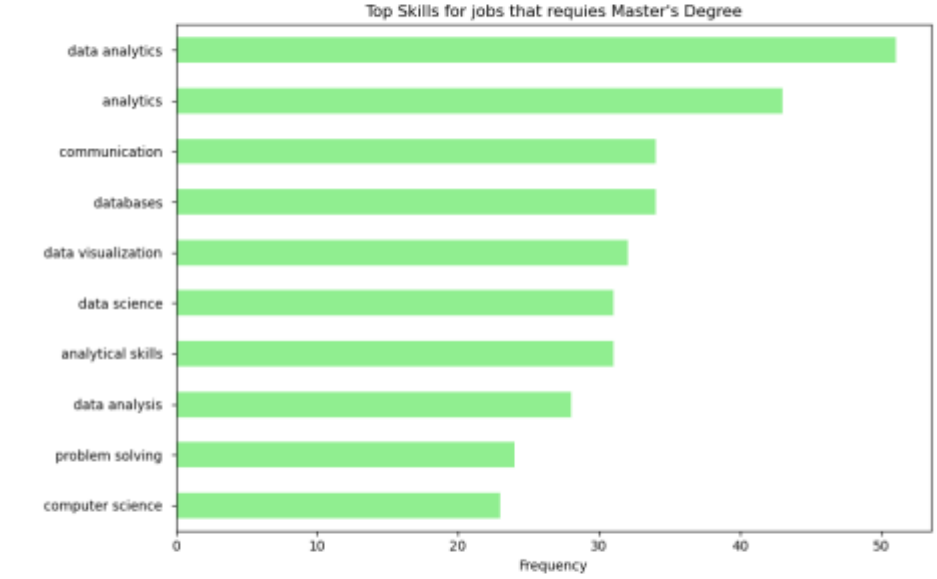
Educational Requirements

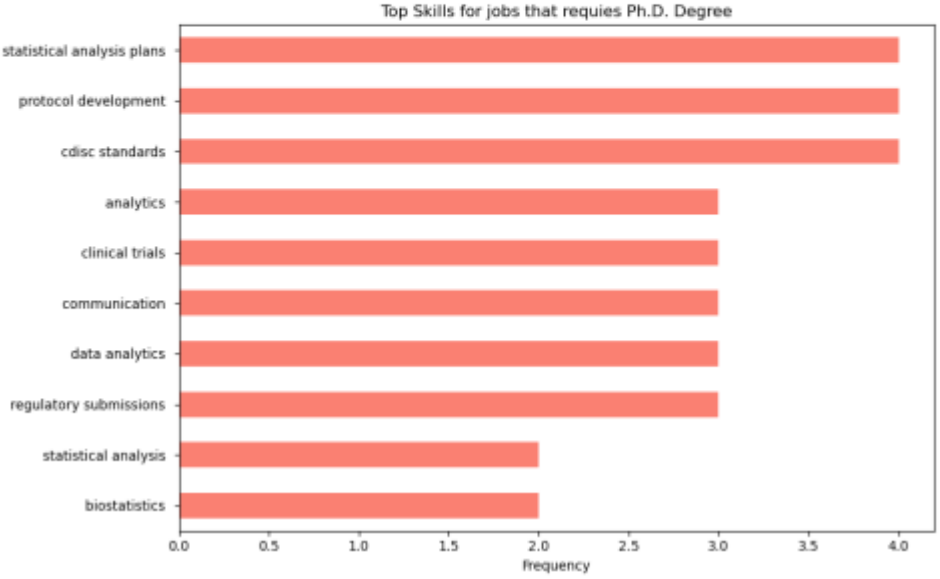
- **Degree Distribution:** Analyzed and visualized the distribution of required educational qualifications.



- **Top Skills by Degree:** Identified top 10 skills for different educational qualifications.


Top Skills for jobs that requies Bachelor's Degree
data analytics
problem solving
communication
analytics
business requirements
data analysis
databases
analytical skills
data visualization
data science
0
20
40
60
80
100
120
140
Frequency


Top Skills for jobs that requies Master's Degree
data analytics
analytics
communication
databases
data visualization
data science
analytical skills
data analysis
problem solving
computer science
0
10
20
30
40
50
Frequency

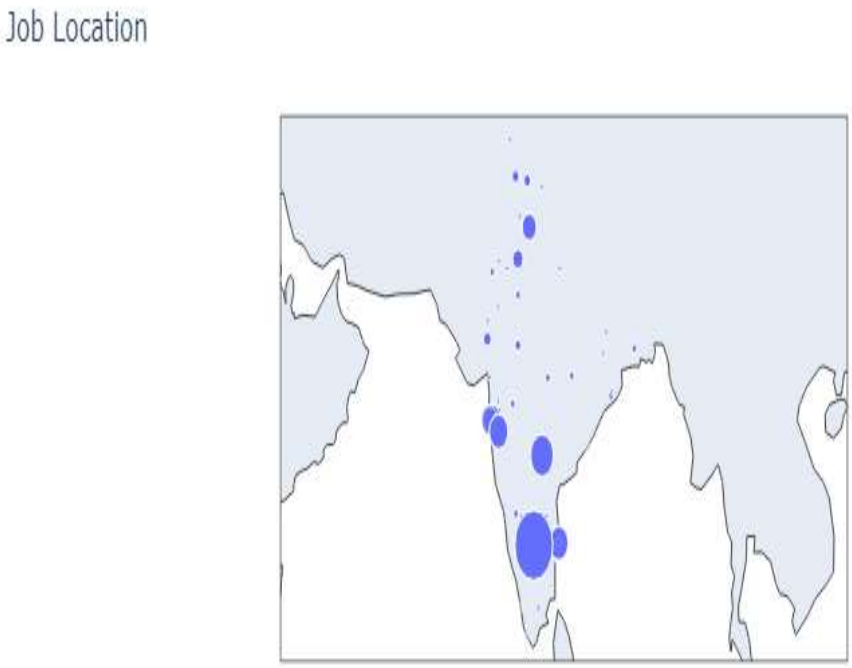

Top Skills for jobs that requies Ph.D. Degree
statistical analysis plans
protocol development
cdisc standards
analytics
clinical trials
communication
data analytics
regulatory submissions
statistical analysis
biostatistics
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Frequency

Work Mode Analysis

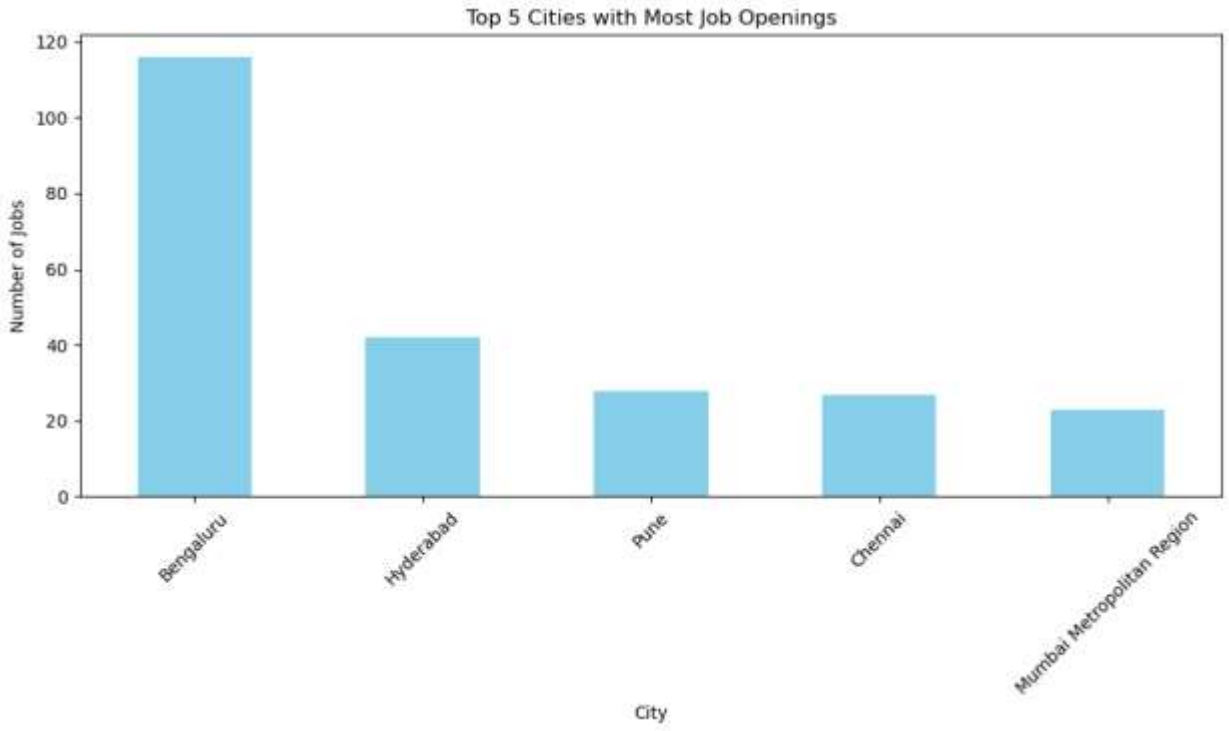
- **Work Mode Distribution:** Analyzed and visualized the distribution of job modes (On-site, Remote, Hybrid).
- **Top Skills by Work Mode:** Identified top 5 skills for different work modes.

Job Location Analysis

- **Job Locations:** Analyzed geographic distribution using city coordinates.

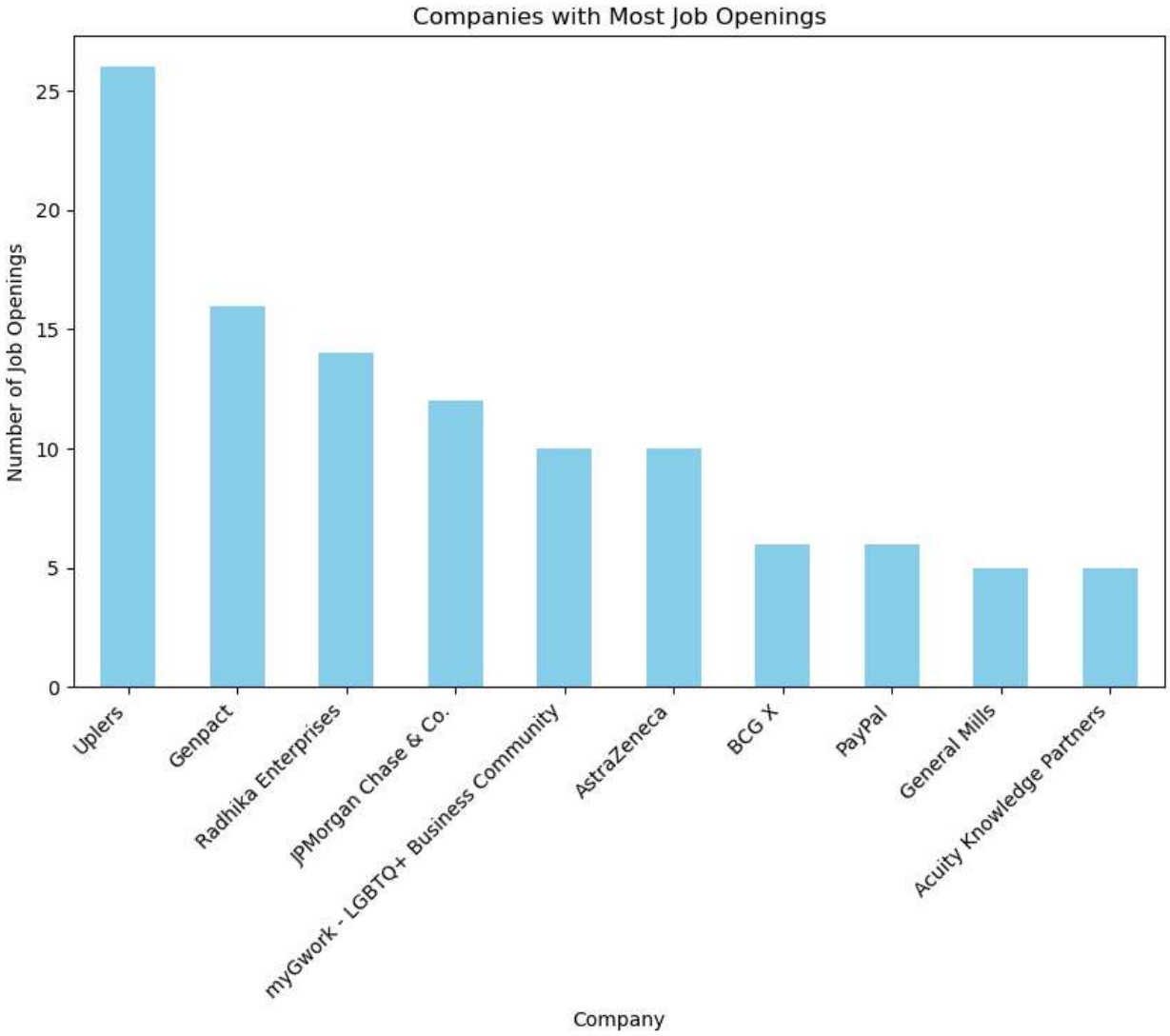


- **Top Cities:** Identified and visualized the top 5 cities with most job openings.

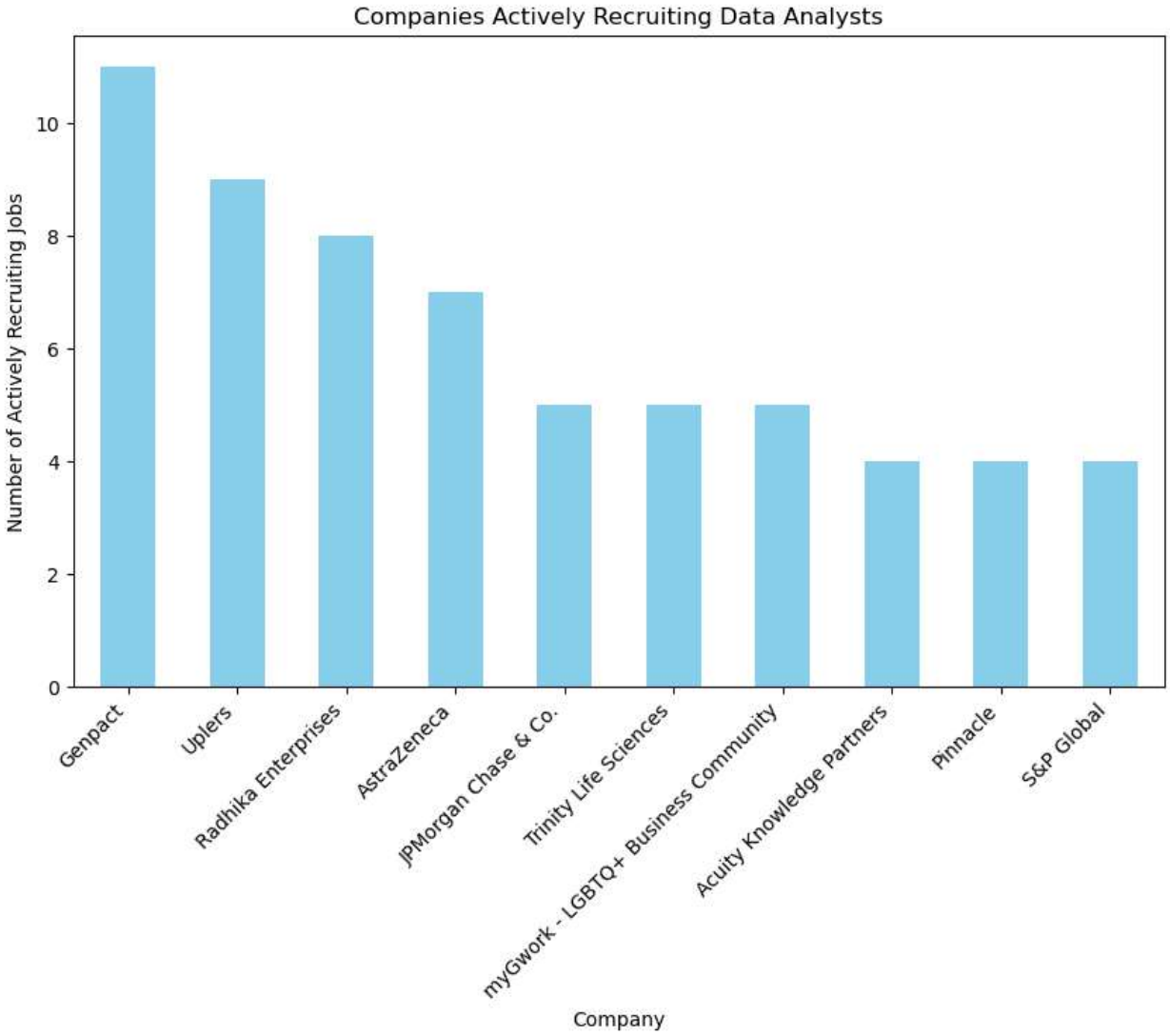


Company Analysis

- **Companies with Most Job Openings:** Identified top 10 companies with the most job openings.

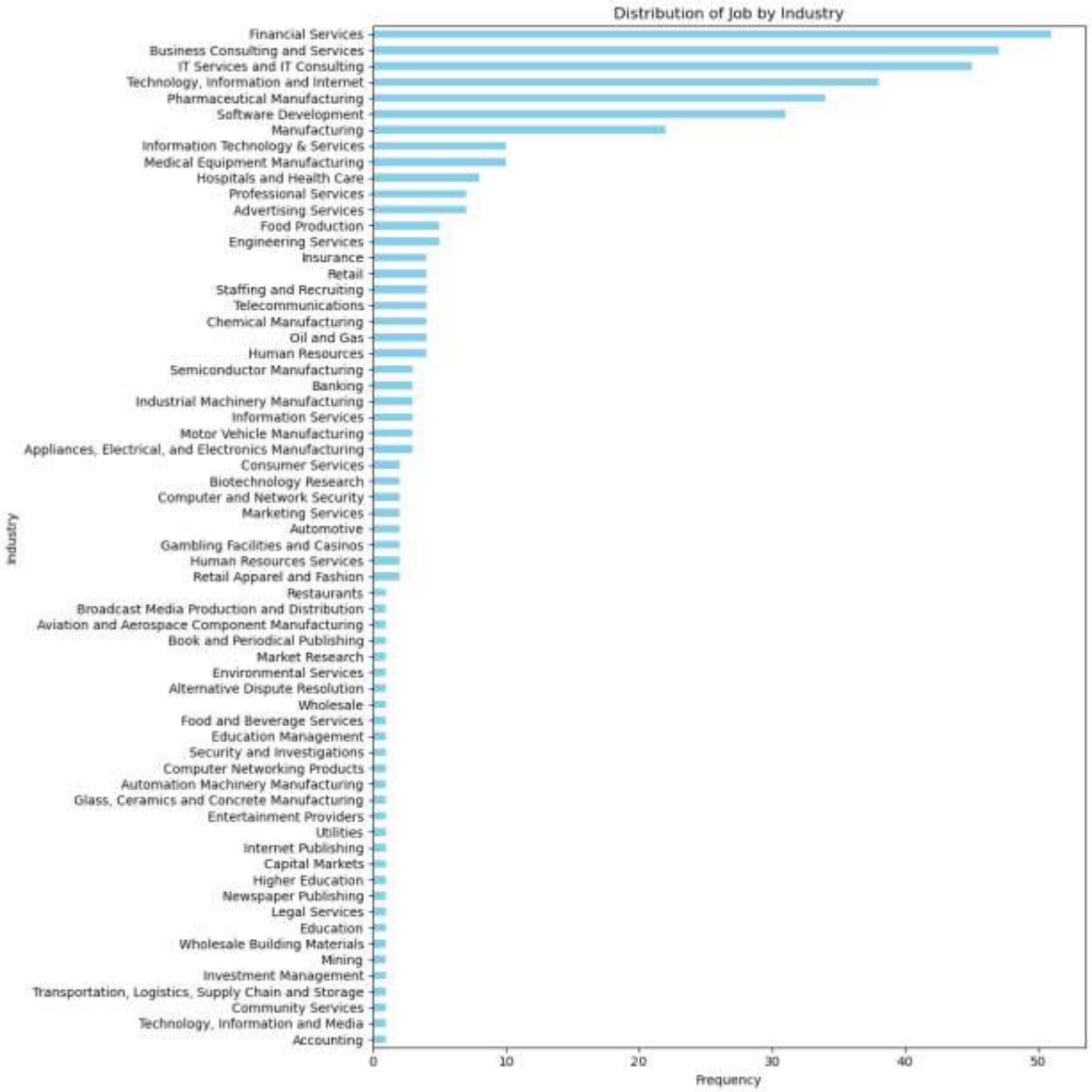


- **Actively Recruiting Companies:** Visualized companies actively recruiting based on job openings.

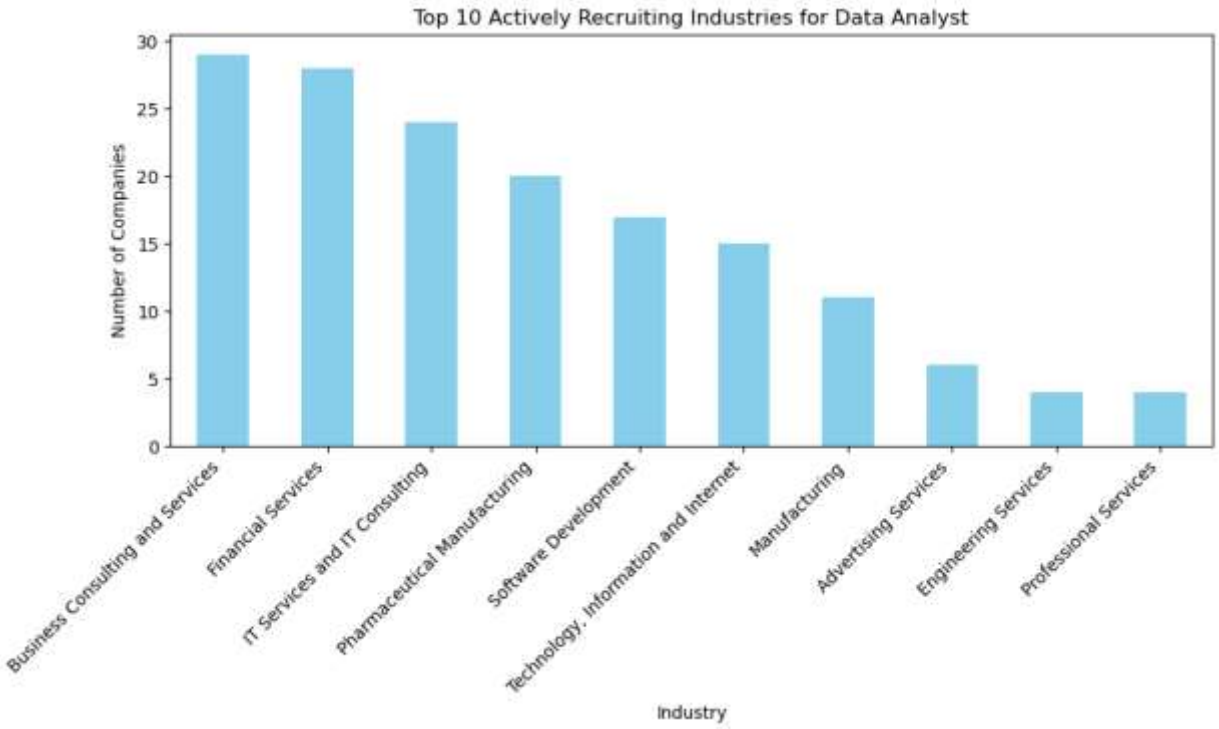


Industry Analysis

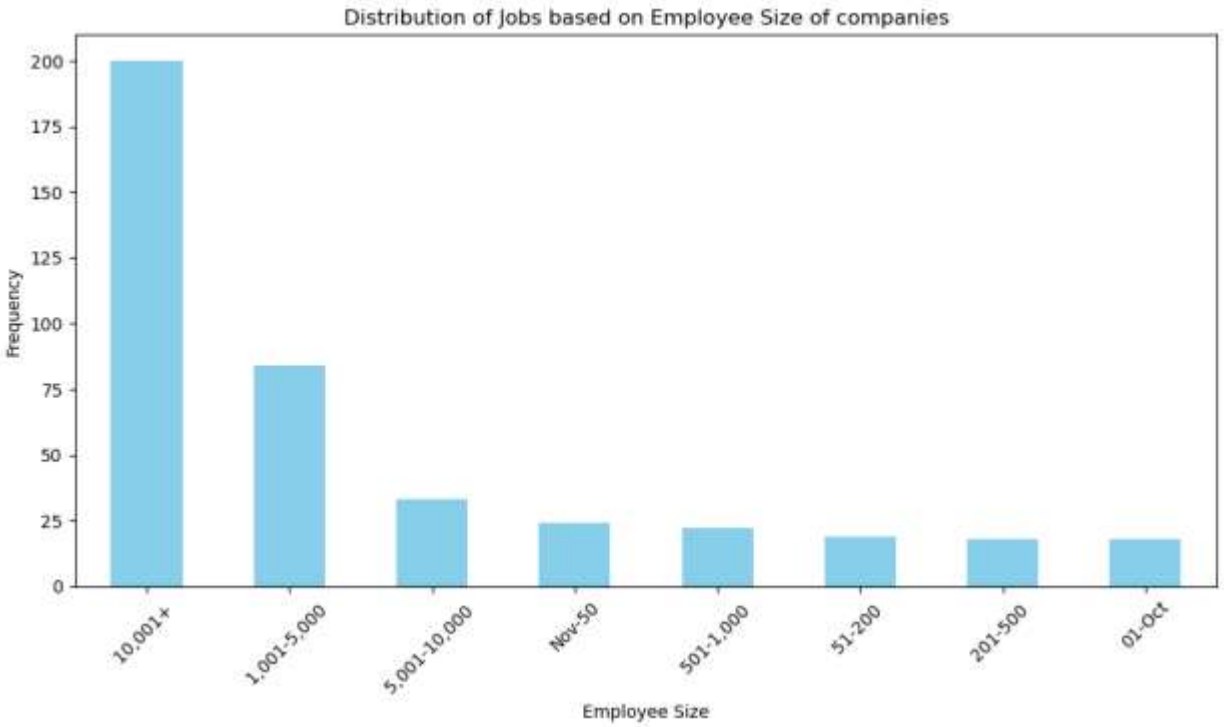
- **Industry Distribution:** Analyzed and visualized distribution of job openings across industries.



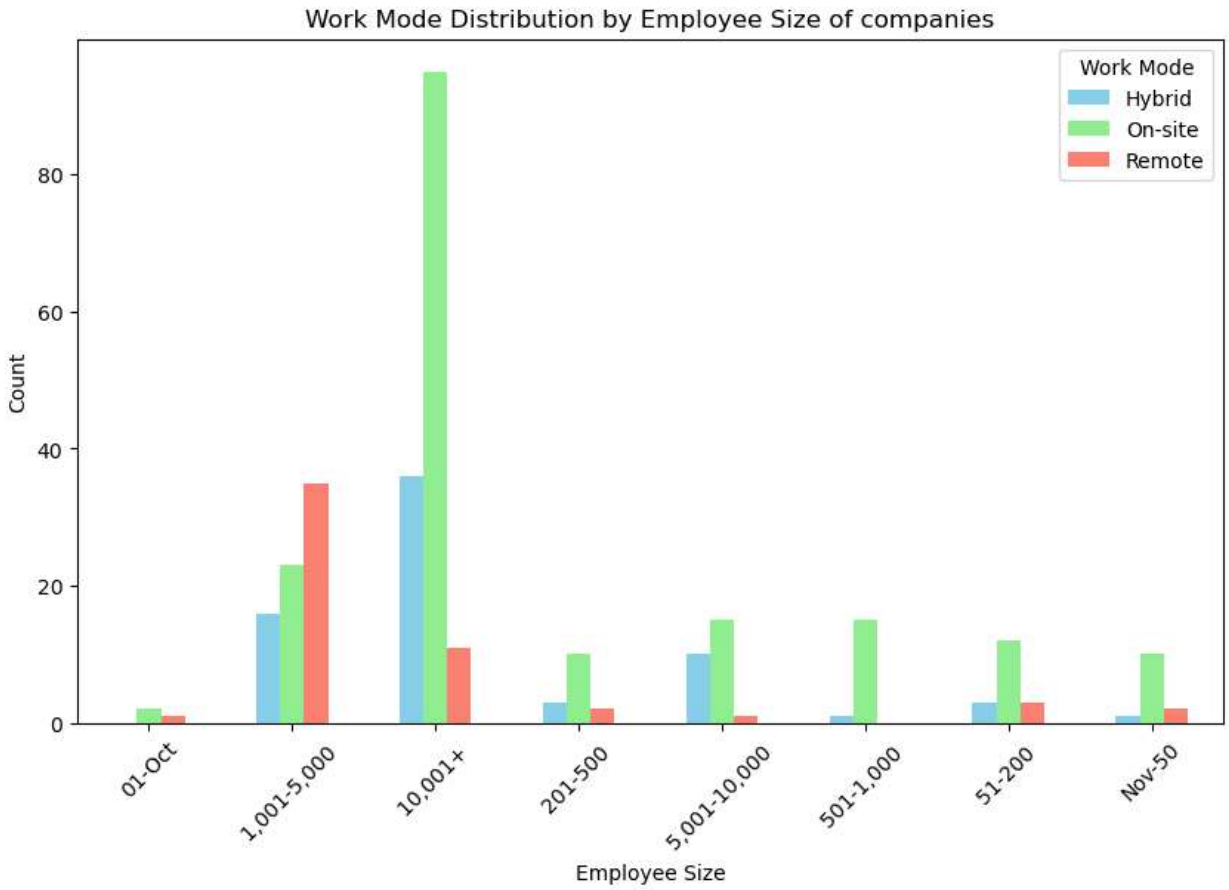
- **Actively Recruiting Industries:** Identified top industries actively recruiting.



- **Employee Size Distribution:** Analyzed job openings based on employee size.



- **Work Mode by Employee Size:** Visualized distribution of work modes across different employee sizes



Hypothesis Testing

Hypothesis 1: Recruitment Status by Industry and Location

- **Null Hypothesis (H0):** Recruitment status depends on Industry and Location.
- **Alternative Hypothesis (H1):** Recruitment status does not depend on Industry and Location.
- **Results:**
    - **Industry:** Chi-square statistic = 95.53, p-value = 0.494. Fail to reject H0.
    - **Location:** Chi-square statistic = 93.50, p-value = 0.379. Fail to reject H0.
    - **Conclusion:** Recruitment status depends on Industry and Location.

Hypothesis 2: Required Experience/Degree among Industries

- **Null Hypothesis (H0):** Required experience or degree is the same across industries.
- **Alternative Hypothesis (H1):** Required experience or degree differs between industries.

- **Results:**
    - **Required Experience:** Chi-square statistic = 578.20, p-value = 0.0001. Reject H0.
    - **Degree:** Chi-square statistic = 388.41, p-value = 7.11e-05. Reject H0.
    - **Conclusion:** Required experience and degrees differ significantly between industries.

CONCLUSION

Key Findings

- **Skills Analysis:** Top skills and skill clusters for Data Analyst roles were identified.
- **Educational Requirements:** Distribution of required educational qualifications was analyzed.
- **Work Mode Preferences:** Distribution of work modes and top skills by work mode were analyzed.
- **Job Location Insights:** Top cities and actively recruiting locations were identified.
- **Company and Industry Trends:** Top companies and industries actively recruiting were analyzed.

Implications for Job Seekers

- Align skills with the most frequently mentioned and clustered skills.
- Understand the preferred work modes and top cities for job openings.
- Target companies and industries with the most job openings and active recruitment.

Recommendations

For Job Seekers:

- Focus on industries and locations where there is active recruitment.
- Tailor your resume to highlight required experiences and degrees that are in demand within your target industry.

For Employers:

- Use data insights to target recruitment efforts in specific locations and industries.
- Ensure job listings highlight the specific experience and educational qualifications that are most sought after.

For Educational Institutions:

- Align curriculum and training programs with the skills and degrees that are in high demand in the job market.

RESOURCES

- **Data Sources:** LinkedIn job postings.
- **Tools and Libraries:** Selenium, Pandas, Numpy, Scikit-learn, NLTK, Matplotlib, Seaborn.