

# Job Market Analysis Report

**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.**

## **Data collection method (include the sample size & process of data collection)**

This report is prepared using the data from indeed job website on the **position of Data Scientist in India**.

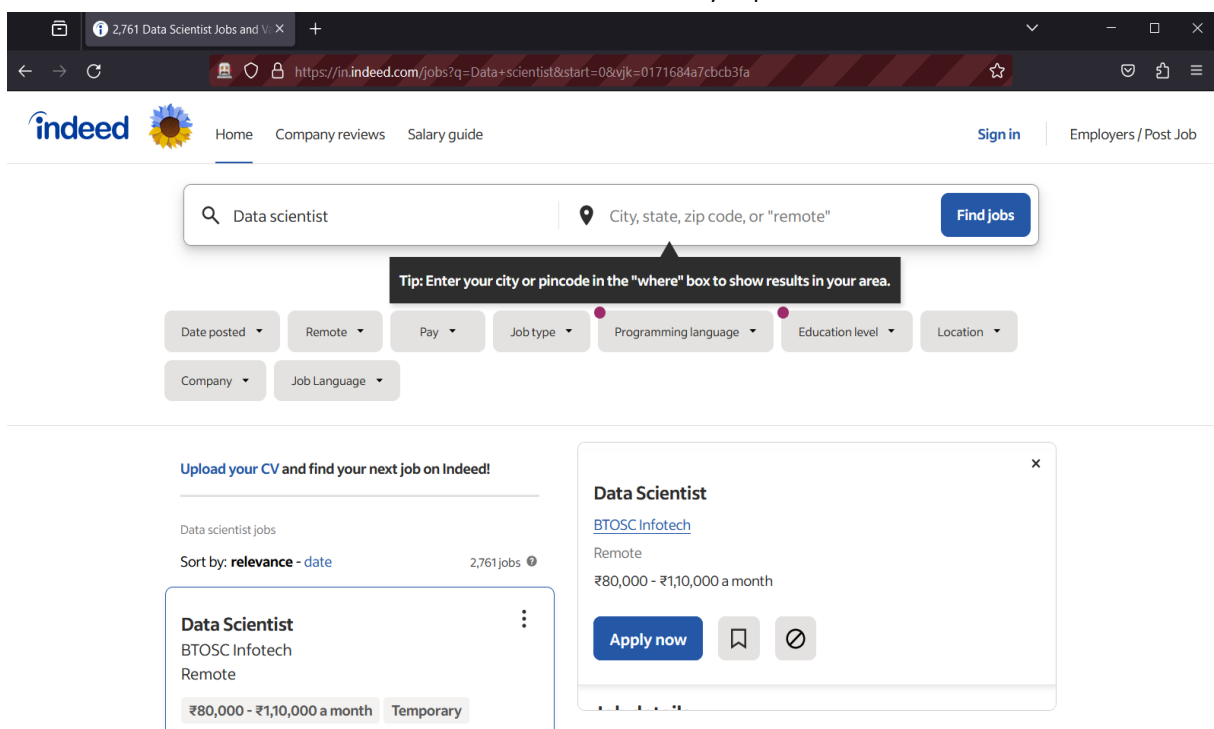
The **total population of the data Jobs** for the above specifications is **approximately around 2782**. The **best sample size is 10% of the population**. Hence, the **sample size for this analysis report** is approximately taken as **300 jobs**, which is 10% of 2782.

As a first step, the Beautiful Soup was chosen as the web scrapping tool to scrape the data from the indeed job website. **The biggest challenge in this step is that the Beautiful Soup was not able to access the indeed website**. Further **analysis revealed the cause as the block provided for the website security by third party firewall** similar to the security services provided by website security company like Cloudflare. This refrain using of the web scrapping tools of Beautiful Soup and Scrappy.

**Alternatively, identified the way to scrap the data using Python Selenium for accessing the HTML Elements present in the certain page and data extracted from the Indeed job website.**

The steps involved in the process of Python Selenium automation to scrap are as follows:

- The Selenium needs to initialize the web driver, in order to launch the browser. In this report Firefox web driver is used.
- The URL needs to be defined for the web driver to navigate to the particular website.
- With the help of the driver elements can be found. Although there are multiple ways to identify elements like Id, Xpath, CSS elements, class, partial text etc. This report uses Xpath for that purpose.
- Store the found element into the dictionary, to convert the stored dictionary to the pandas dataframe.
- Save the dataframe into the csv file for further data analysis process.



*Image 1: Sample of the python Selenium Web driver*

## Market Data Visualization

This step involves in the exploration of the data scraped from the Indeed Job Website.

The collected data frame is displayed below.

	Job_title	company	Location	salary	contract	links	Full_Description
0	Data Scientist	PayU	Hybrid remote in Bengaluru, Karnataka	Negotiable	Full-time	<a href="https://in.indeed.com/rc/clk?jk=05dadf1768f24f...">https://in.indeed.com/rc/clk?jk=05dadf1768f24f...</a>	About PayU\nPayU, a leading payment and Fintec...
1	Data Scientist - Matching Scienc	Indeed	Remote	₹32,00,000 - ₹48,00,000 a year	None	<a href="https://in.indeed.com/company/Indeed/jobs/Data...">https://in.indeed.com/company/Indeed/jobs/Data...</a>	Our Mission\nAs the world's number 1 job site*...
2	Data Scientist	Discover Dollar	Bengaluru, Karnataka	Negotiable	Full-time	<a href="https://in.indeed.com/rc/clk?jk=55474e4a2eac11...">https://in.indeed.com/rc/clk?jk=55474e4a2eac11...</a>	Job Description:\n\nWe are looking for a Data ...

Image 2: Sample of Scraped data represented in the form of Dataframe

The data consists of 300 rows and 7 columns. The data type of the 7 columns are object variables.

Column	Description
Job_title	Type of job title that the recruiter has posted on the Indeed Website.
Company	Name of the company.
Location	Location of the company.
Salary	Salary that the recruiter is ready to offer to the candidate.
Contract	The candidate is working for full time or part time.
Links	To access the job posting
Full_Description	Briefing on the job details that the recruiter wants about the specific role.

There are 98 duplicate rows present in the data and removed for data cleaning and precise analysis.

**Due to removal of the duplicate rows now the sample size is reduced to 202 rows.**

There are **no missing values present in the data.**

Deep diving into the analysis with the graphical interpretation of the data.

### Job Title:

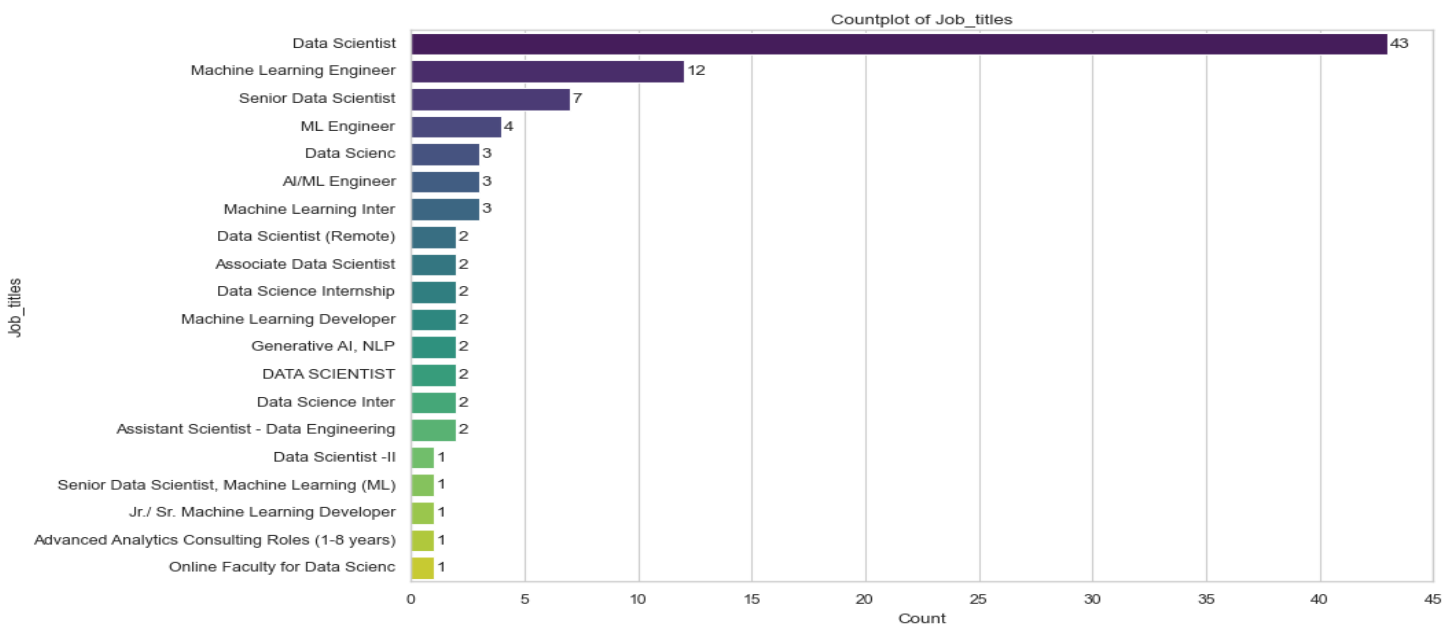


Image 4: Count plot of the job titles

The above is the graphical representation of the few of the various job titles present in the sample.

There is a total of **126 different job title** in the projected sample.

The top 5 list of most common job titles are shown below:

- Data Scientist – 43
- Machine Learning Engineering – 12
- Senior Data Scientist – 7
- ML Engineer – 4
- Data Science – 3

### Location:

There are a **total of 40 locations of the company** present from the projected sample.

Here, the location is defined with two further details as city and state. There are **17 states** and **40 cities** present of the 40 locations of our sample.

**Bengaluru is the city where most of the job openings are available.** Since, Bengaluru is in Karnataka, reveals the fact that **the most job openings are from Karnataka state.**

**Next ranks the remote jobs** that shows the feasibility of the work location and can visit the office only during necessity.

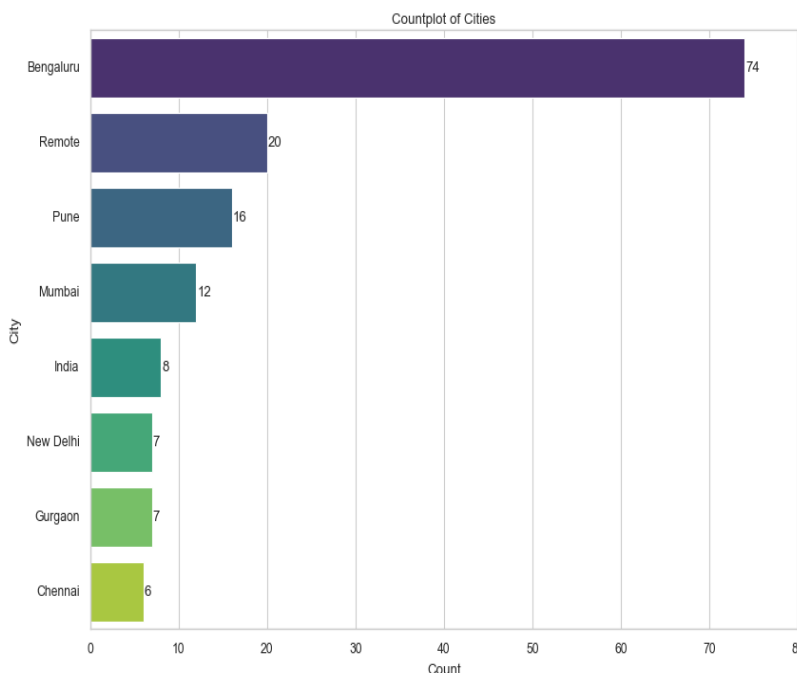


Image 5: Top 5 city job posted

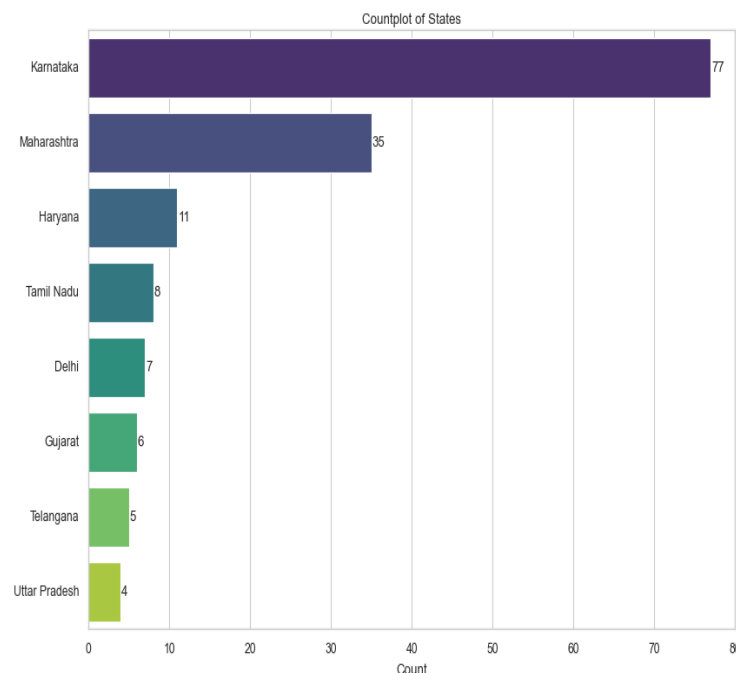


Image 6: Top 5 state job posted

### Types of Position:

This parameter explains whether the job is full-time or Part-time.

**Full-time:** Employee works for a predetermined number of hours per week, mostly around 35 to 40 hours, and is typically eligible for benefits offered by the employer.

**Part-time:** Means employee works fewer hours than the standard full-time schedule, mostly with reduced benefits.

**None:** Meaning it can be either Full-time or Part-time without any mention.

The below pie chart represents the None and Full-Time.

**There are 136 None Values contract jobs and 66 Full-time contract jobs.**

The projected sample represents that **most of the jobs are Full-Time** as **None type are** can be either **Full/Part-time**.

There are only **38 Part-Time jobs** in the indeed job website for the total population of 2782 Jobs.

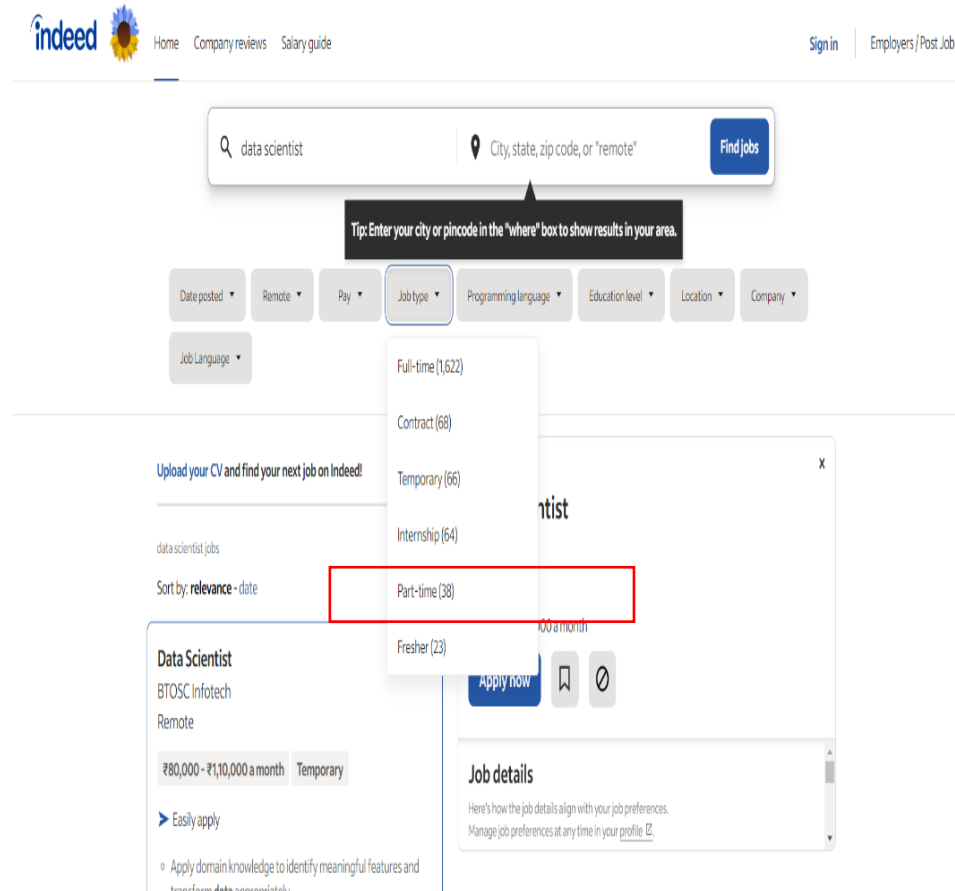
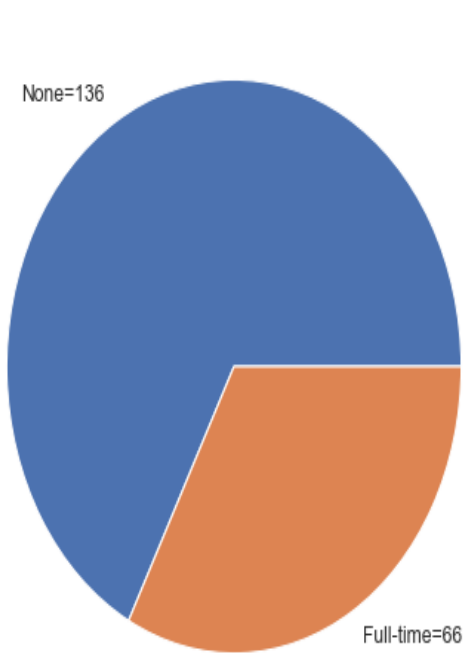


Image 7: Pie chart for None type and Full-Time

Image 8: Number of Part-Time for total population

### Key Skills:



Image 9: List of words from the description

The Key Skills required for this position is described in the job post as **descriptions**.

All these descriptions are stored in the list as full description.

This list dataframe is passed to the wordcloud to generate a plot with **all major important words highlighted as bigger font size** and **minor important word in smaller font size**.

The top 7 Major Key skill are as follows:

- Machine Learning
- Experience
- Python
- Data science
- Model
- Team
- Solution

### Your ideal job:

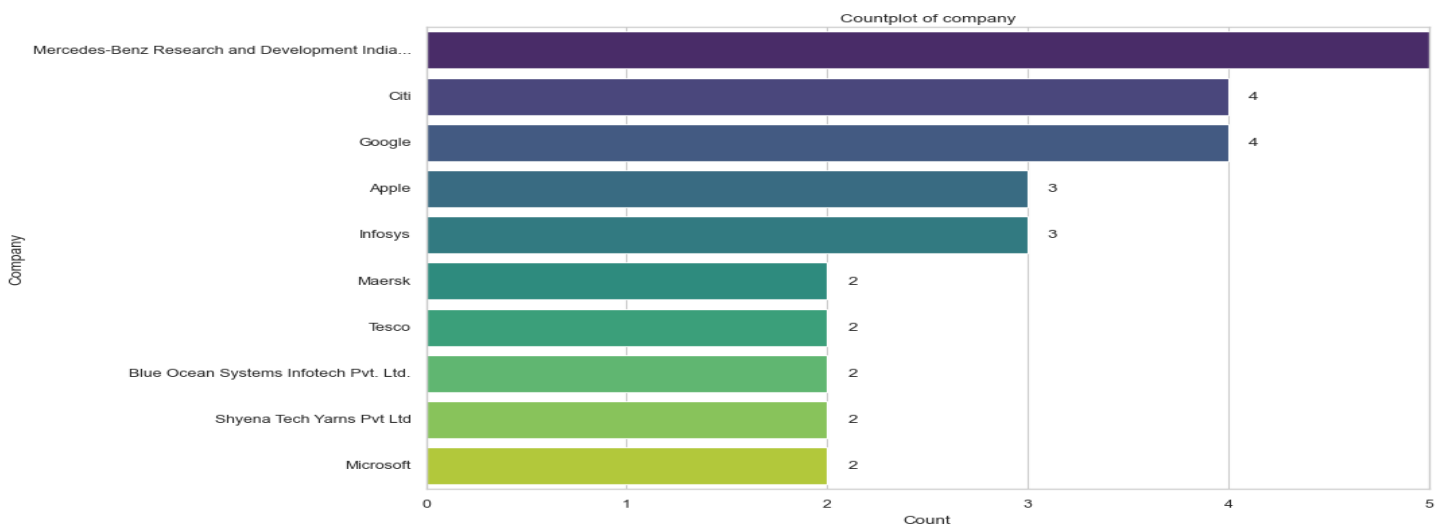
**Data scientist** is considered as the ideal job for me.

The top relevant key skills required for it are as follows:

- Programming Language (Python or R programming)
- Statistical Analysis
- Data Mining
- Machine Learning
- Database Expertise (SQL)
- Deep learning
- Artificial Intelligence (AI)
- Big Data

## Miscellaneous Analysis with the data:

### Company:



*Image 10: Top 10 Companies that hire most in our sample*

This field represents the company which post the job on the Indeed Job Website.

The top 5 companies that are posting job for Data Science are:

- Mercedes-Benz Research and Development India

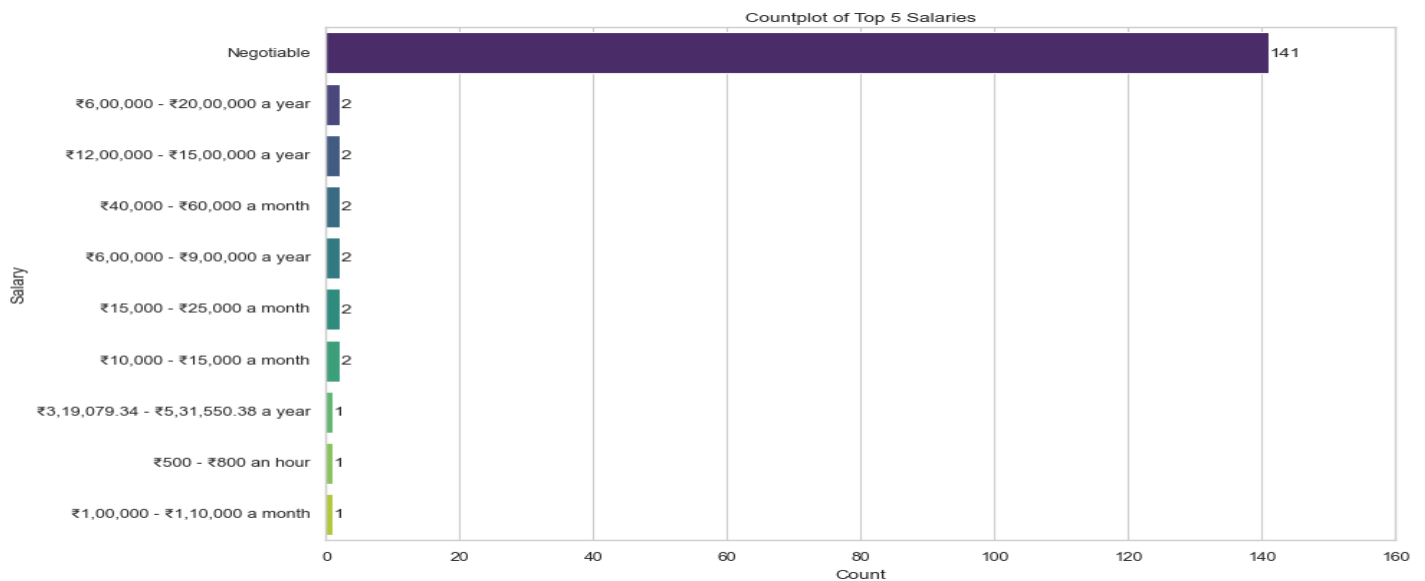
- Citi
- Google
- Apple
- Infosys

The FAANG Companies like Apple and Google are hiring for 3 jobs and 4 jobs respectively.

### Salary:

This field represents the Salary that recruiter of the job wants to offer to the applicant.

The report identifies that **majority of the recruiters are ready to do negotiation** on the salary for their job posts.



*Image 11: Top 10 Salary fields for the Job posts*