

Naukri Data Analyst Job Analysis

1. Introduction

In this project, we downloaded the **Naukri Data Analyst Jobs dataset** from **Kaggle** and performed data cleaning, skill extraction, and basic exploratory analysis to understand job trends, top hiring locations, and in-demand skills.

2. Scraping Logic (as per dataset creation)

The dataset originally collected Naukri job listings using the following scraping logic:

- Cleaning text fields with `.text.strip()`
- Extracting structured fields such as:
 - **Job Role, Company, Job Type**
 - **Location, Experience, Salary**
 - **Description, Key Skills**
 - **Posted Date, Apply Link**

This logic resulted in a structured dataset ready for cleaning and analysis.

3. Data Cleaning Steps

We cleaned and standardized the dataset using:

✓ Text Cleaning

- Removed spaces using `.str.strip()`
- Standardized text to lowercase

✓ Salary Cleaning

- Removed symbols & commas
- Extracted numeric salary using **regex**
- Created `Cleaned_Salary` column

✓ Experience Cleaning

- Extracted years using regex
- Created Experience_Years

✓ Key Skills Extraction

- Split using regex: `[,/|•]`
- Standardized skills to lowercase
- Counted skill frequency using **Counter**

4. Basic Analysis

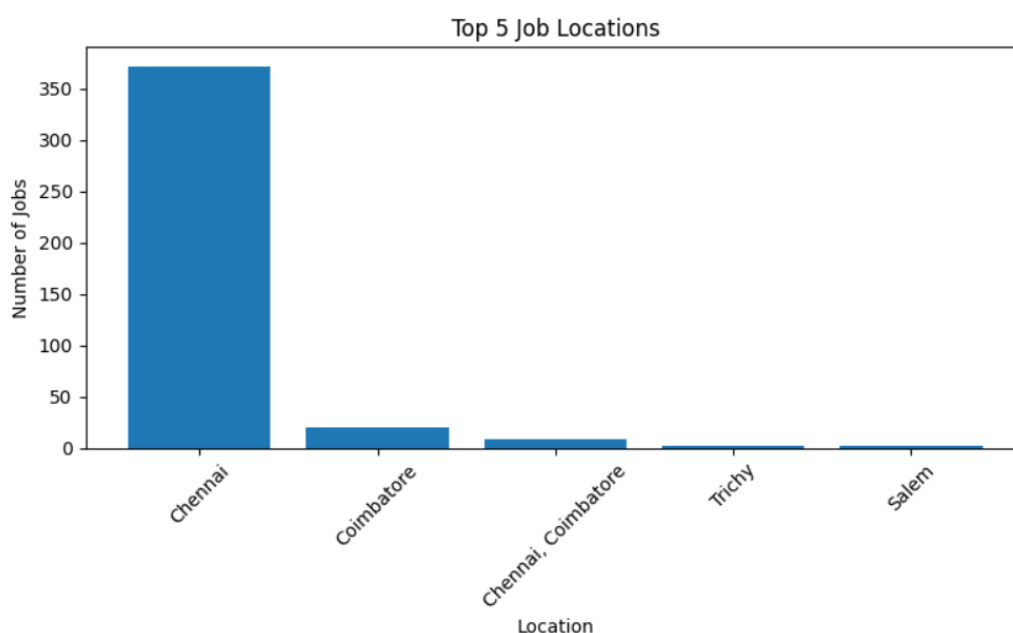
Total Jobs Scraped

Total jobs: $\rightarrow \text{len(df)}$

Output: 409

Top Job Locations

We identified the **Top 5 hiring locations** using `.value_counts()` and visualized them using a bar chart.



Most In-Demand Skills

By extracting and counting skills, the most common ones appeared to be:

```
[('metadata data management test scripts talent acquisition process improvement pharma sas programming manager quality control',
 4),
('data analysis sql queries analytical pivots ms excel communication analytical skills excel',
 2),
('revenue cycle management us healthcare advanced ms excel data analyst microsoft office suite microsoft excel query communication',
 2),
('quantum data processing spss communication ascii excel mrdcl', 2),
('data analysis image processing renewable energy autocad analytical microstation remote sensing analytics',
 2),
('sap di docker ci cd python sap data intelligence nodejs qa data sources',
 2),
('it skills quality assurance cloud digital marketing data analysis change management manager quality assurance direct marketing',
 2),
('business system analyst warehouse translationproduct portfolio diversity and inclusion agile methodology information technology data warehousing',
 2),
('data analysis tableau data science diversity and inclusion cloud relationship programming information technology',
 2),
('azure ml analytical jupyter azure verbal written communication sql tensorflow',
 2)]
```

5. Challenges Faced

During scraping and dataset generation, the following challenges were common:

- Dynamic job content loading
 - Inconsistent salary & experience formats
 - Skills listed in mixed formats (comma, slash, bullets)
 - Missing or incomplete fields
 - Anti-bot detection requiring delays
-

6. Conclusion

Using the Kaggle-sourced Naukri dataset, we processed and analyzed Data Analyst job listings through data cleaning, regex-based extraction, and visual exploration. The analysis highlights major hiring locations and the skills most demanded in the Indian job market for Data Analyst roles.