

Job Market Analysis: Web Scraping, SQL Data Cleaning, and Power BI Dashboarding

This project analyzes the job market for **Data Analyst roles in India** by scraping job listings from LinkedIn. The extracted data includes **job title, company name, location, salary, and required skills**.

Project Workflow:

- 1 Web Scraping** – Automated data extraction using **Selenium** in Jupyter Notebook.
- 2 SQL Data Cleaning** – Removing duplicates, handling missing values, and structuring data.
- 3 Power BI Dashboard** – Visualizing key insights like salary distribution, top hiring companies, and required skills.

◆ Data Cleaning SQL Queries

✍ 1. Remove Duplicates

```
DELETE FROM job_listings
WHERE id NOT IN (
    SELECT MIN(id) FROM job_listings GROUP BY title, company, location
);
```

✍ 2. Standardize Job Titles (Lowercase & Trim)

```
UPDATE job_listings
SET title = TRIM(LOWER(title));
```

✍ 3. Standardize Company Names (Fix case issues)

```
UPDATE job_listings
SET company = INITCAP(company); -- Capitalizes first letter of each word
```

✍ 4. Handle Missing Locations (Replace NULL with 'Remote' if applicable)

```
UPDATE job_listings
SET location = 'Remote'
WHERE location IS NULL OR location = '';
```

Insights from SQL Queries

Using SQL, we had derived valuable insights from the job market data:

SQL Queries for Job Market Analysis

Easy Queries (Basic Insights)

- 1 Total number of job listings**
- 2 Top 5 locations with the most job openings**

- 3 Top 5 companies hiring the most Data Analysts**
- 4 Average salary for Data Analyst roles**

Medium Queries (Intermediate Insights)

- 5 Salary distribution by experience level**
- 6 Most in-demand skills for Data Analysts**
- 7 Salary comparison for remote vs. on-site jobs**
- 8 Companies offering the highest salaries**

Hard Queries (Advanced Insights)

- 9 Finding Skills Most Frequently Mentioned in Job Titles**
- 10 Analyzing Job Role Diversity in Companies**

```
-- sql queries

-- Count Total Job Listings
select count(*) jobs from job_listings;

-- Top 5 Companies Hiring the Most
select company , count(*) as job_count
from job_listings
group by Company
order by job_count desc
limit 5;

-- Most Common Job Titles
select title , count(*) as count
from job_listings
group by title
order by count desc
limit 6;

-- Count of Remote Jobs
select count(*) as number_of_remote_jobs
from job_listings

-- Count of Remote Jobs
select count(*) as number_of_remote_jobs
from job_listings
where location = "remote";

-- Percentage of Remote vs. On-Site Jobs
SELECT
    (COUNT(CASE WHEN LOWER(location) LIKE '%remote%' THEN 1 END) * 100.0 / COUNT(*)) AS remote_percentage,
    (COUNT(CASE WHEN LOWER(location) NOT LIKE '%remote%' THEN 1 END) * 100.0 / COUNT(*)) AS onsite_percentage
FROM job_listings;

-- Top 5 Locations with the Most Job Openings
select location , count(*) as count
from job_listings
group by location
order by count desc
limit 5;
```

```
-- Average Word Count in Job Titles (To Identify Role Trends)
SELECT AVG(LENGTH(title) - LENGTH(REPLACE(title, ' ', '')) + 1) AS avg_word_count
FROM job_listings;

-- Find Companies Posting Jobs for Multiple Locations
SELECT company, COUNT(DISTINCT location) AS location_count
FROM job_listings
GROUP BY company
HAVING location_count > 1
ORDER BY location_count DESC
LIMIT 5;

-- Ranking Companies Based on Their Share of Job Listings
SELECT company, COUNT(*) AS job_count,
       RANK() OVER (ORDER BY COUNT(*) DESC) AS ranking
FROM job_listings
GROUP BY company
LIMIT 10;
```

```
-- Finding Skills Most Frequently Mentioned in Job Titles
SELECT
    skill, COUNT(*) AS mention_count
FROM (
    SELECT CASE
        WHEN LOWER(title) LIKE '%sql%' THEN 'SQL'
        WHEN LOWER(title) LIKE '%python%' THEN 'Python'
        WHEN LOWER(title) LIKE '%excel%' THEN 'Excel'
        WHEN LOWER(title) LIKE '%power bi%' THEN 'Power BI'
        WHEN LOWER(title) LIKE '%tableau%' THEN 'Tableau'
        WHEN LOWER(title) LIKE '%r%' THEN 'R'
        WHEN LOWER(title) LIKE '%aws%' THEN 'AWS'
        WHEN LOWER(title) LIKE '%cloud%' THEN 'Cloud'
        ELSE 'Other'
    END AS skill
    FROM job_listings
) AS skill_table
WHERE skill != 'Other'
GROUP BY skill
ORDER BY mention_count DESC;

-- Analyzing Job Role Diversity in Companies
SELECT company, COUNT(DISTINCT title) AS unique_roles
FROM job_listings
GROUP BY company
HAVING unique_roles > 1
ORDER BY unique_roles DESC
LIMIT 10;
```

Insights for Power BI Dashboard

Using Power BI, we can create visualizations to present these insights:

Key Dashboard Components

-  **Job Market Overview:** Total job listings, top hiring companies, and top locations.
-  **Salary Analysis:** Average salary by experience level, company-wise salary comparison.
-  **Skill Demand:** Top 10 most in-demand skills for data analysts.
-  **Job Posting Trends:** Monthly growth in job postings over time

