

From Raw Data to Strategic Insight: Decoding the Job Market

- **What:** An end-to-end analytics dashboard built to transform 54,000+ raw job postings from Naukri.com into an interactive strategic tool.
- **Why:** The Indian job market is complex and fast-moving. Actionable intelligence on skills, salaries, and roles is critical for job seekers, recruiters, and businesses.
- **How:** By implementing a robust ETL pipeline and data model, the project delivers clear, data-driven answers on hiring trends, skill demand, and salary benchmarks.



SQL DATABASE



POWER BI VISUALIZATION

54,000+

Job Postings Analyzed

The Challenge: Unstructured Data Hides Critical Insights

- **Fragmented Sources:** Job data is inherently inconsistent and lacks a unified structure, making direct analysis unreliable and difficult.

“3-5 yrs”
“5-10 LPA” MySQL
“5-10 LPA Postgresql
skill_1 “dirty data”
Not Disclosed
Bangalore/Bengaluru **skill_2**
skill_8
3.20-3.40LPA

- **Low Visibility:** Key decision-making data—like competitive salary benchmarks and in-demand skills—is buried within unstructured text.

- **Schema Flaws:** Raw data often violates fundamental database principles (e.g., skills stored across 8 separate columns), preventing accurate analysis of skill demand.

The Foundation: A Snapshot of the Indian Tech Job Market



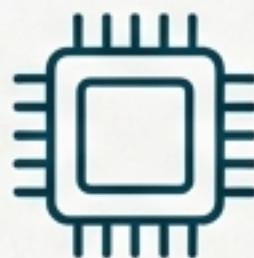
Source: Naukri.com



Records: 54,000+
job postings



Date: March 2023

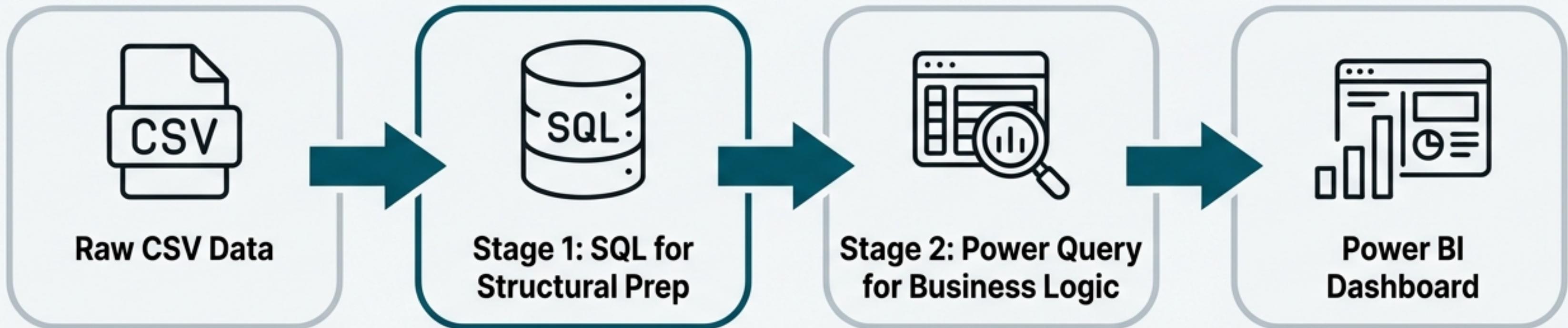


Focus: Data, Analytics,
and IT roles in India

Key Data Challenges

- **Text-Based Metrics:** Salary and experience were stored as text ranges (e.g., “3.20-3.40 LPA,” “3-5 yr”), requiring parsing for quantitative analysis.
- **Wide Data Format:** Skills were spread across multiple columns (`skill_1` to `skill_8`), making it impossible to aggregate or trend skill demand effectively.
- **Inconsistent Naming:** The same skills (e.g., SQL variants), locations, and roles appeared under multiple different names.

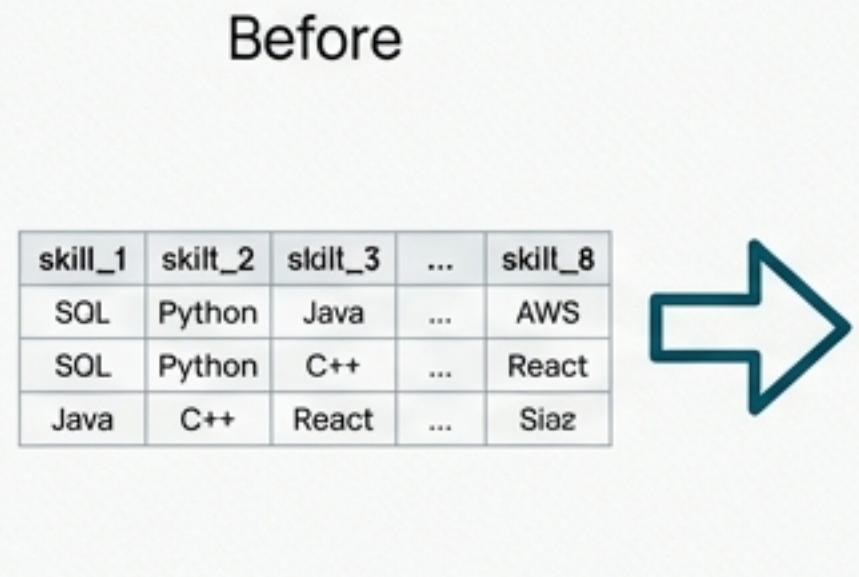
The Strategy: A Decoupled, Two-Stage Transformation Pipeline



- **Stage 1: SQL for Structural Foundation:** Used a MySQL database for heavy-lifting operations, including initial ingestion and schema normalization (unpivoting).
- **Stage 2: Power Query for Business Logic:** Leveraged Power Query for fine-grained cleaning, text parsing, and applying flexible business rules.
- **Why This Architecture?:** Decoupling the data layer (SQL) from the BI layer (Power BI) creates a more scalable, maintainable, and efficient solution. SQL handles bulk transformations, while Power Query excels at iterative refinement.

From Messy to Meaningful: Core Data Transformations

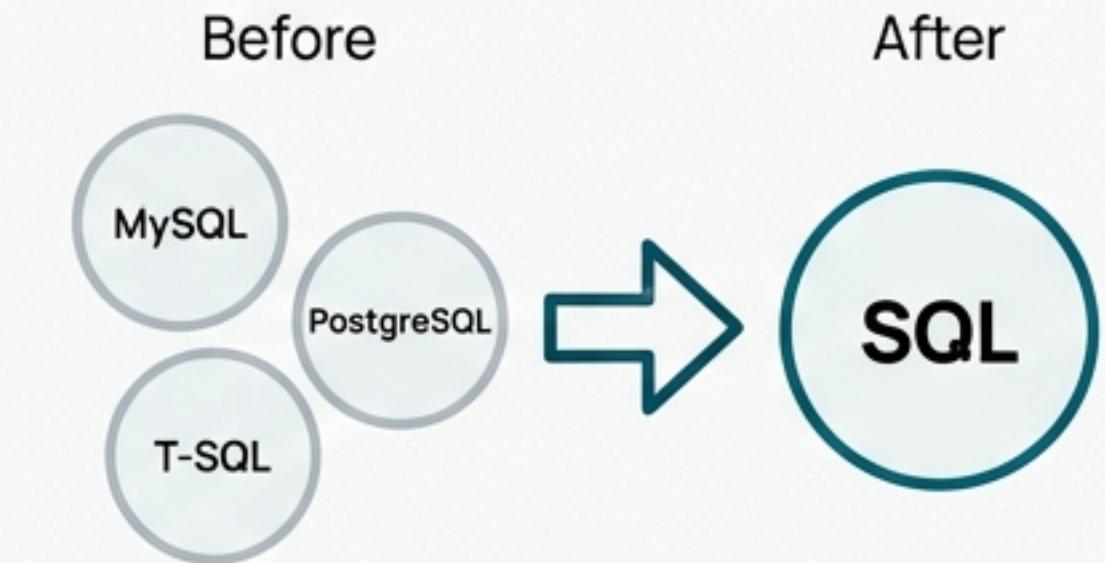
Skill Unpivoting (Wide to Long)



Salary & Experience Parsing



Skill Standardization

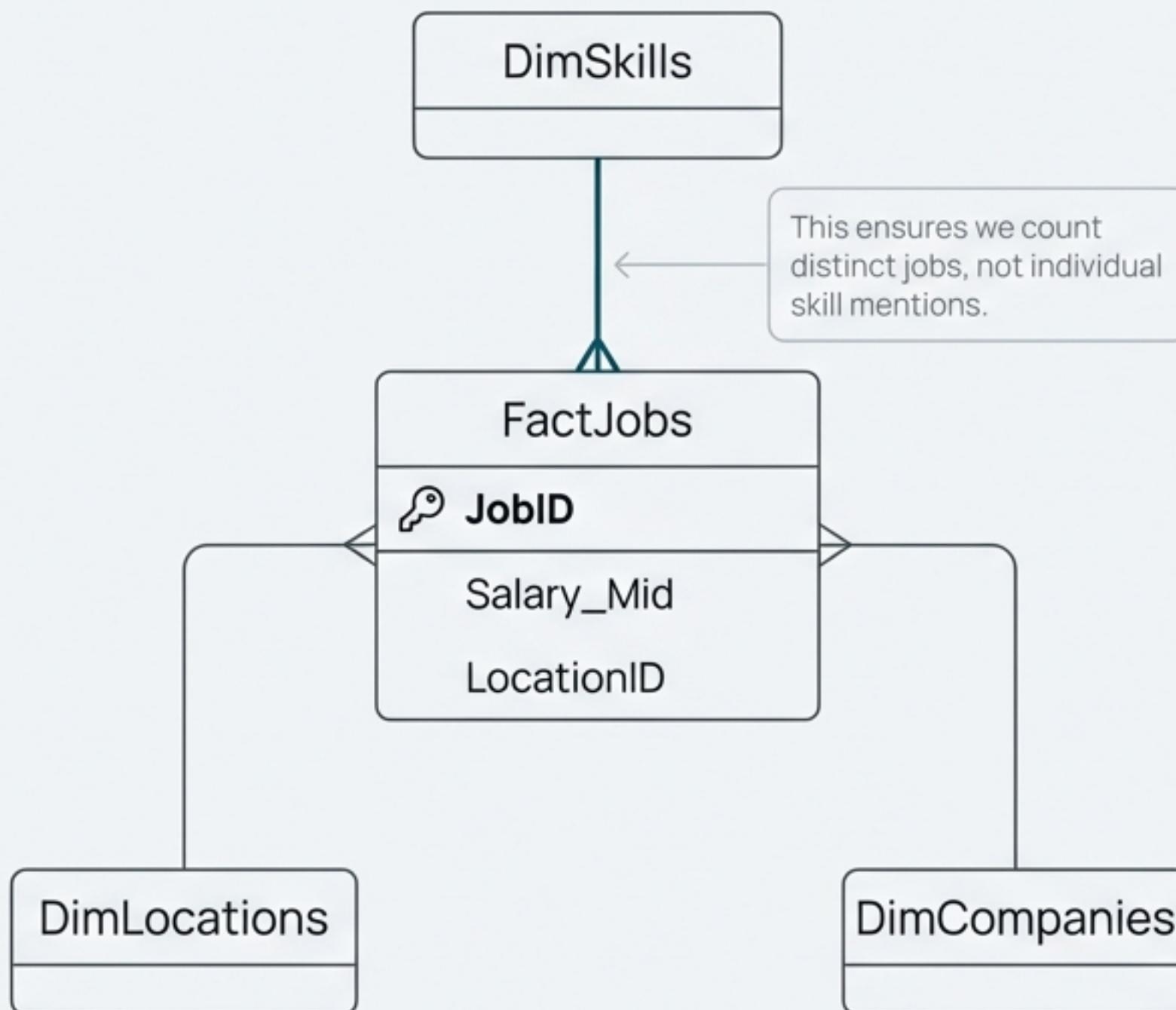


- Transformed 8 separate skill columns into a single, normalized skill column, solving the '1NF violation' and enabling proper skill analysis.

- Developed logic to extract numerical midpoints from text ranges (e.g., '3.20-3.40 LPA' became '3.3'), making salaries and experience levels quantifiable.

- Created grouping logic to consolidate variations (e.g., 'MySQL', 'PostgreSQL', 'T-SQL') into a single 'SQL' category, preventing fragmented insights.

The Blueprint: A Data Model Built for Analytical Integrity



- **Granularity Defined:** The model is centered on a single fact: the “Job Posting.” Each job is assigned a unique ID to serve as the primary grain.
- **One-to-Many Relationships:** A clean one-to-many relationship was established between the `Jobs` table and the `Skills` table.
- **Preventing Fan-Out Errors:** This design is critical. It ensures that when filtering by a skill, we count the number of *distinct jobs*, not the total number of skills mentioned, avoiding double-counting and guaranteeing accurate reporting.

Powering Insights: An Intelligent DAX Calculation Engine

```
Total Job Postings =  
DISTINCTCOUNT(FactJobs[JobID])
```



Ensures we count jobs, not rows of skills, maintaining analytical integrity.

```
High Pay % =  
DIVIDE(  
CALCULATE(  
[Total Job Postings],  
FactJobs[Salary_Mid] > 12  
),  
[Total Job Postings])
```



Creates a robust, context-aware KPI that is safe from division-by-zero errors and responds correctly to filters.

- **Job-Centric Measures:** Core KPIs like `Total Job Postings` were built using `DISTINCTCOUNT` on the Job ID, directly aligning with the data model's grain.
- **Context-Aware KPIs:** Used functions like `CALCULATE` and `DIVIDE` to build robust, percentage-based metrics (e.g., 'High-Paying Jobs %') that adapt correctly to user filters.
- **Designed for Performance:** The DAX logic was optimized for a smooth, responsive user experience, even with complex filtering.

The User Experience: A Guided Analytical Journey



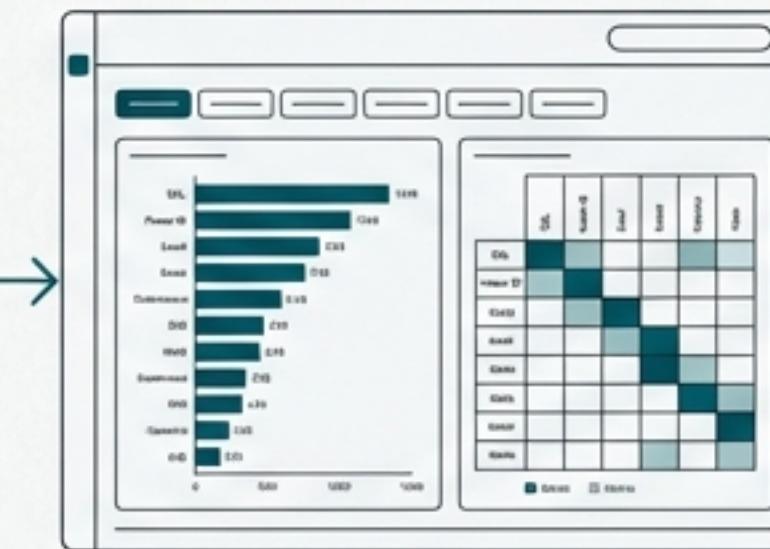
Page 1 - Market Overview

- A high-level summary with headline KPIs: 5.4K total postings, average salary, and hiring trends over time.



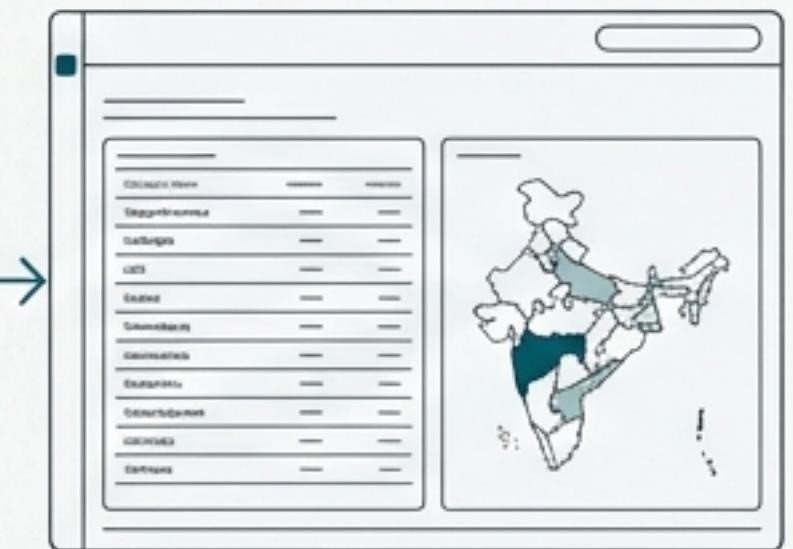
Page 2 - Roles & Experience

- A deep dive into specific job roles, comparing salary benchmarks against required years of experience using scatter plots and bar charts.



Page 3 - Skills Analysis

- An interactive view of the most in-demand skills (SQL, Power BI, Excel), their prevalence across different roles, and their evolution.

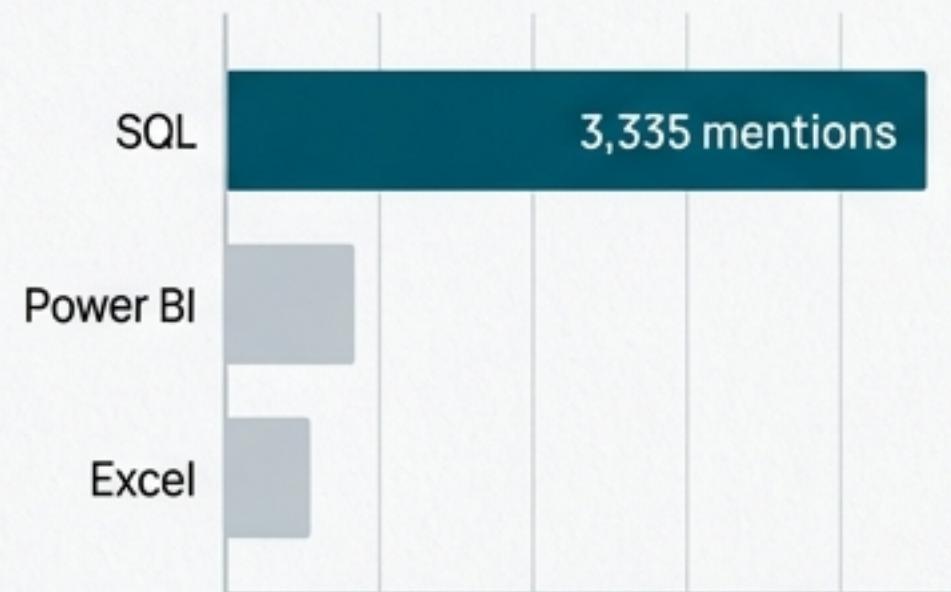


Page 4 - Insights & Companies

- Highlights top hiring companies (e.g., Teleperformance, InfoCepts) and geographical salary differences.

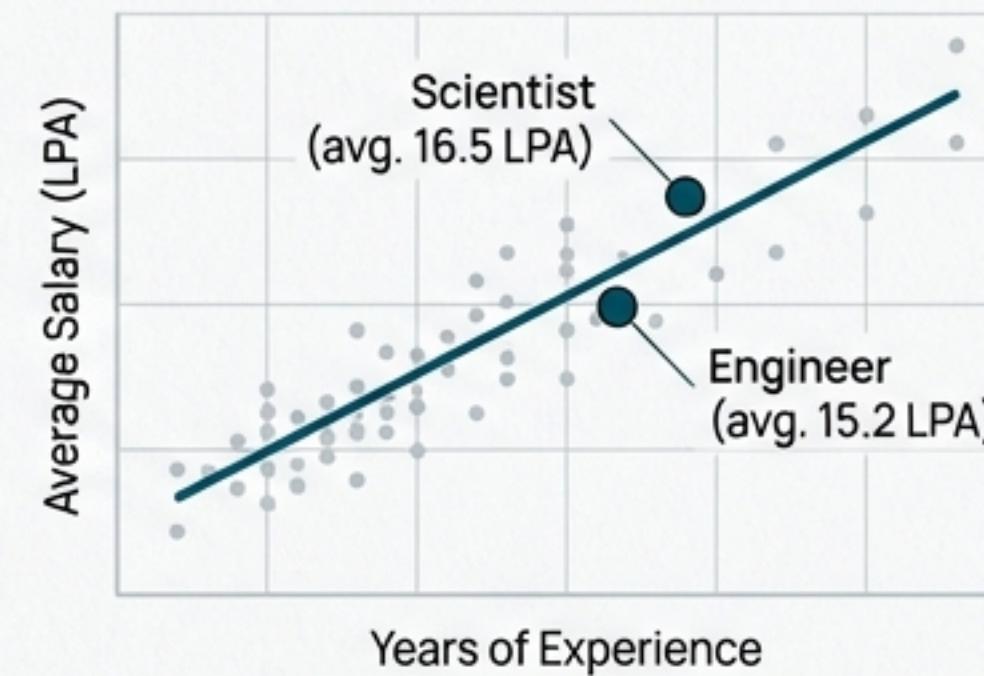
What the Data Revealed: Key Market Insights

SQL Dominance



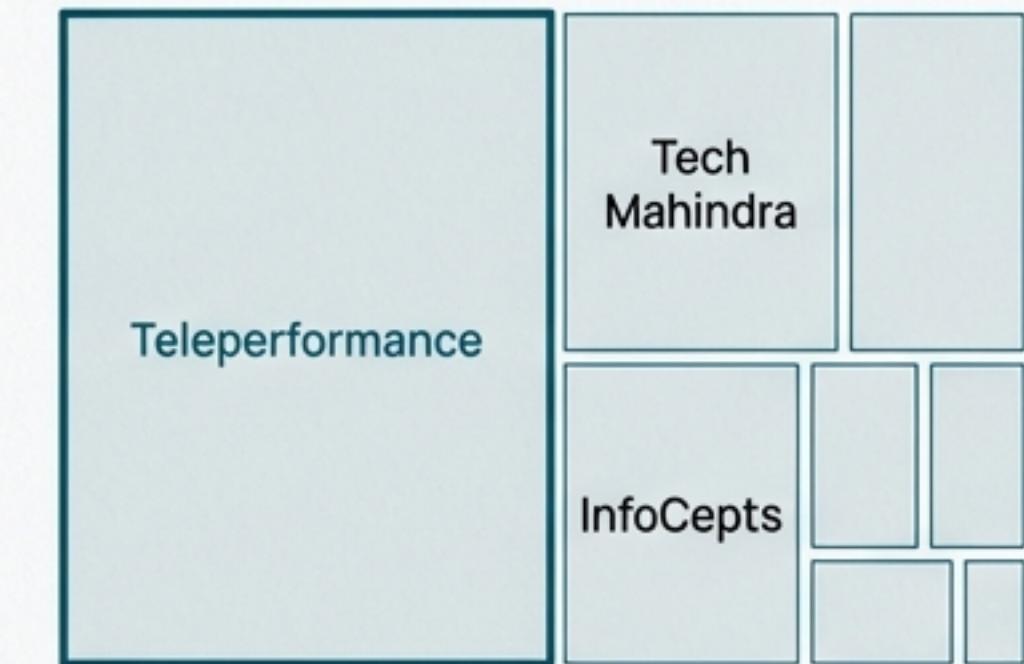
SQL is the Foundational Skill: SQL appeared in the highest percentage of all data-related job postings (3,335 mentions), making it the most critical skill across roles.

Experience & Compensation



Experience Directly Drives Compensation: A clear, positive correlation exists between years of experience and salary, especially for Engineer (avg. 15.2 LPA) and Scientist (avg. 16.5 LPA) roles.

Concentrated Hiring



Hiring is Concentrated: A small number of large companies, such as Teleperformance and Tech Mahindra, account for a significant portion of total job postings.

From Insights to Action: The Business Value of the Dashboard



For Job Seekers

- Provides a data-driven roadmap to identify the most valuable skills to learn and how to benchmark salary expectations.

For Companies & Recruiters

- Enables competitive salary benchmarking, helps define realistic job descriptions, and informs data-driven talent acquisition strategies.

For Analysts

- Serves as a reusable, end-to-end framework for transforming and analyzing similarly unstructured datasets, demonstrating proven best practices.

Core Competencies Demonstrated

-  **End-to-End Project Ownership:** Managed the entire analytics lifecycle, from raw data ingestion and ETL to data modeling, final dashboard delivery, and insight generation.
-  **Technical and Strategic Thinking:** Applied strong data modeling principles (1NF, granularity) and made deliberate technology choices (SQL vs. Power Query) to build a scalable and accurate solution.
-  **Business-First Mindset:** Focused on translating complex technical decisions into clear, actionable business insights that directly answer relevant stakeholder questions.

The Road Ahead: Future Enhancements and Scalability

