Industrial Human Resource Geo-Visualization

Industrial Classification of Main and Marginal Workers: Analysis & Dashboard

1. Project Overview

Background:

The industrial classification of the workforce is crucial for understanding the distribution of labor across sectors. This project aims to update the outdated data on main and marginal workers (excluding cultivators and agricultural laborers) by providing current insights on worker distribution by sex, section, division, and class.

Objective:

To provide policymakers and employment planners with accurate and relevant data by:

Merging multiple datasets.

Analyzing trends using natural language processing.

Visualizing insights through an interactive dashboard.

Scope:

The analysis covers various industries such as manufacturing (plastic, rubber, chemicals, furniture, construction), retail, poultry, agriculture, and more, across different states and demographic groups.

2. Problem Statement

Issue:

Existing data on industrial classification is outdated and fails to reflect current trends in the workforce.

Goal:

Update and analyze the data to reveal:

Distribution of main vs. marginal workers.

Gender-based trends.

Sector-specific insights.

3. Dataset Description

Content:

The dataset contains state-wise counts of main and marginal workers (male and female) across various industrial sectors.

Industries Included:

Manufacturing (e.g., plastic products, rubber products, chemicals, furniture, construction), retail, agriculture, poultry, etc.

4. Approach

Data Pipeline:

Data Collection & Merging:

Merge all provided CSV files into a unified dataframe.

Exploratory Data Analysis (EDA):

Assess data distribution, identify missing values and outliers.

Visualize basic statistics to understand trends.

Data Cleaning:

Handle missing data and outliers.

Standardize industry names and classifications.

Feature Engineering:

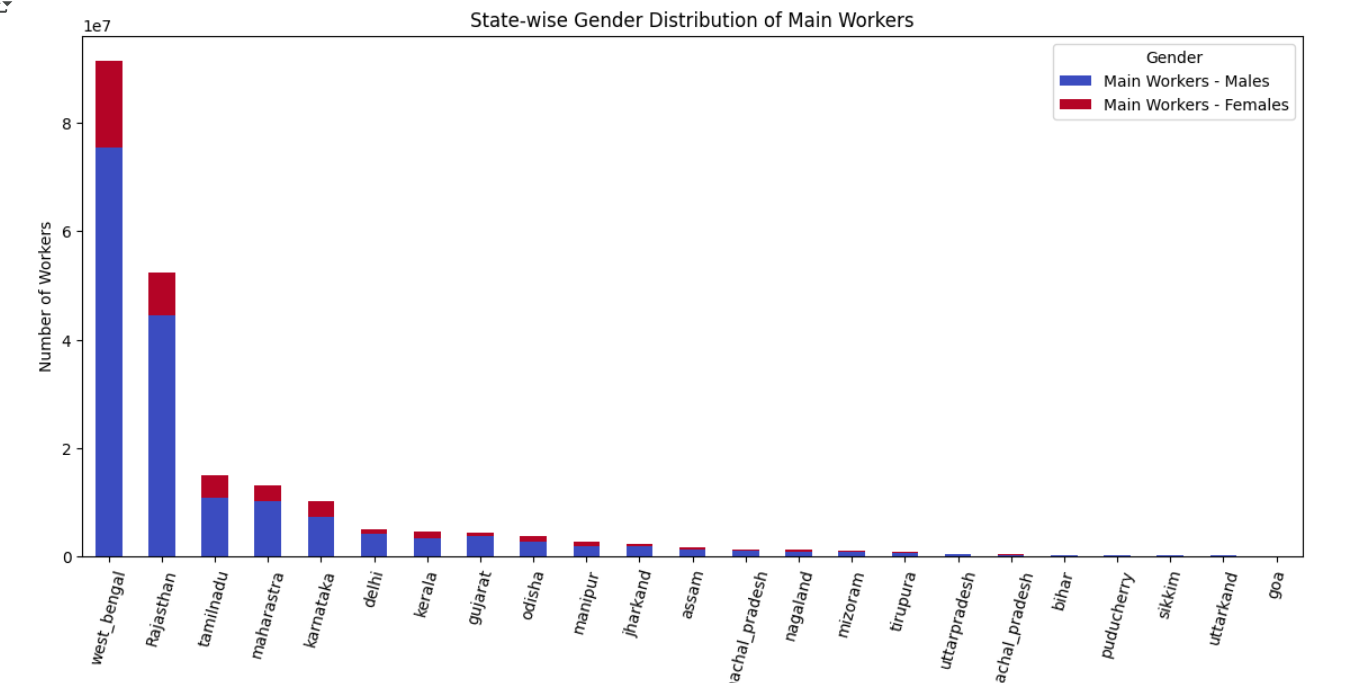
Extract and create meaningful features.

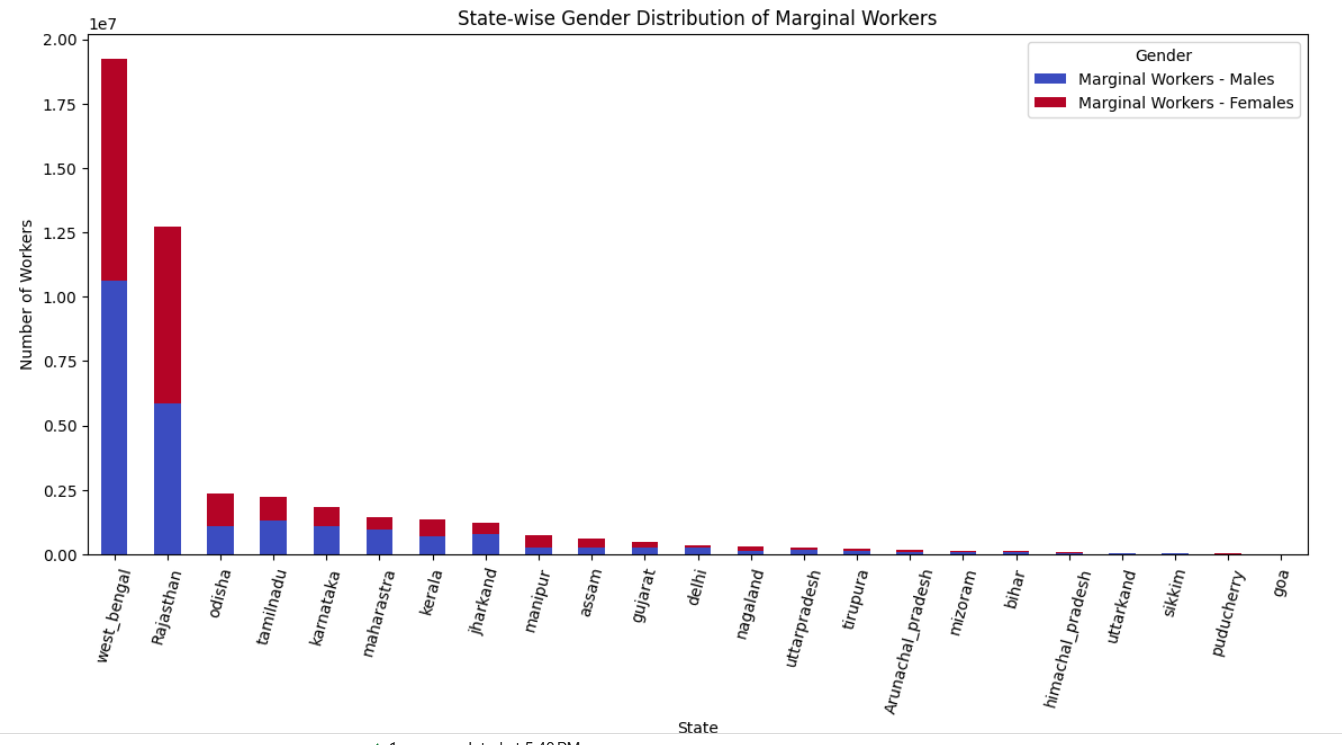
Use NLP techniques to categorize and group core industries (e.g., grouping related business categories like Retail, Poultry, Agriculture, Manufacturing).

Dashboard Development:

Develop an interactive dashboard using Streamlit and Plotly for real-time visualization and analysis.

Insights:

**STATEWISE GENDER DISTRIBUTION OF MAIN WORKERS**

**STATEWISE GENDER DISTRIBUTION OF MARGINAL WORKERS**

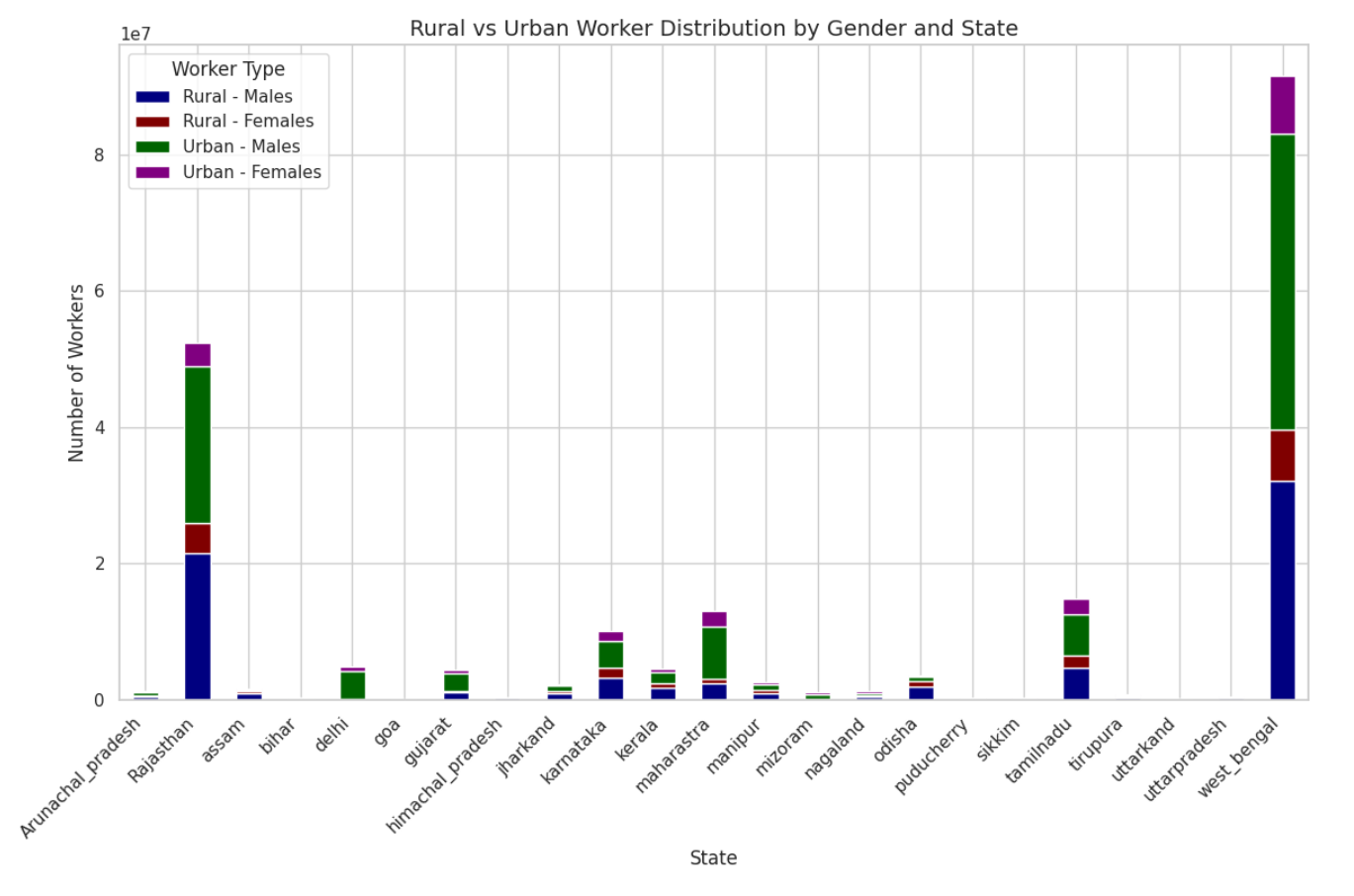
 **Volume:** Main workers significantly outnumber marginal workers across states, indicating more stable employment.

 **Gender Imbalance:** Main workers are predominantly male, while marginal workers sometimes show a more balanced or even female-skewed distribution.

 **Regional Trends:** Southern states tend to have a higher ratio of main to marginal workers, suggesting more stable job markets.

 **Data Consistency:** Discrepancies (e.g., in Manipur) highlight the need for data validation across both employment categories.

**RURAL VS URBAN WORKERS BY STATE AND GENDER**



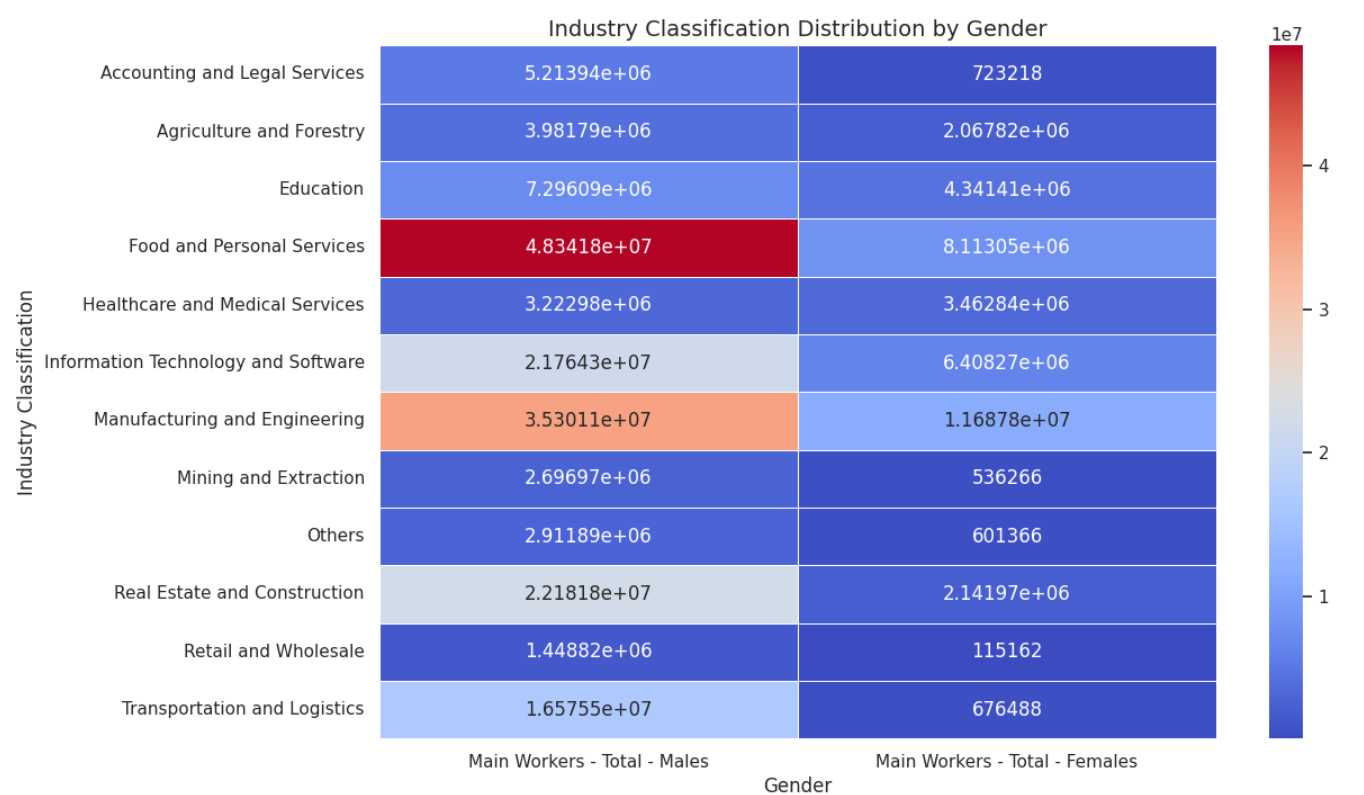
 **Delhi’s Urban Dominance:**  
Delhi’s urban worker numbers (≈4.12 million males and ≈0.69 million females) far exceed its rural counts, underscoring its metropolitan focus.

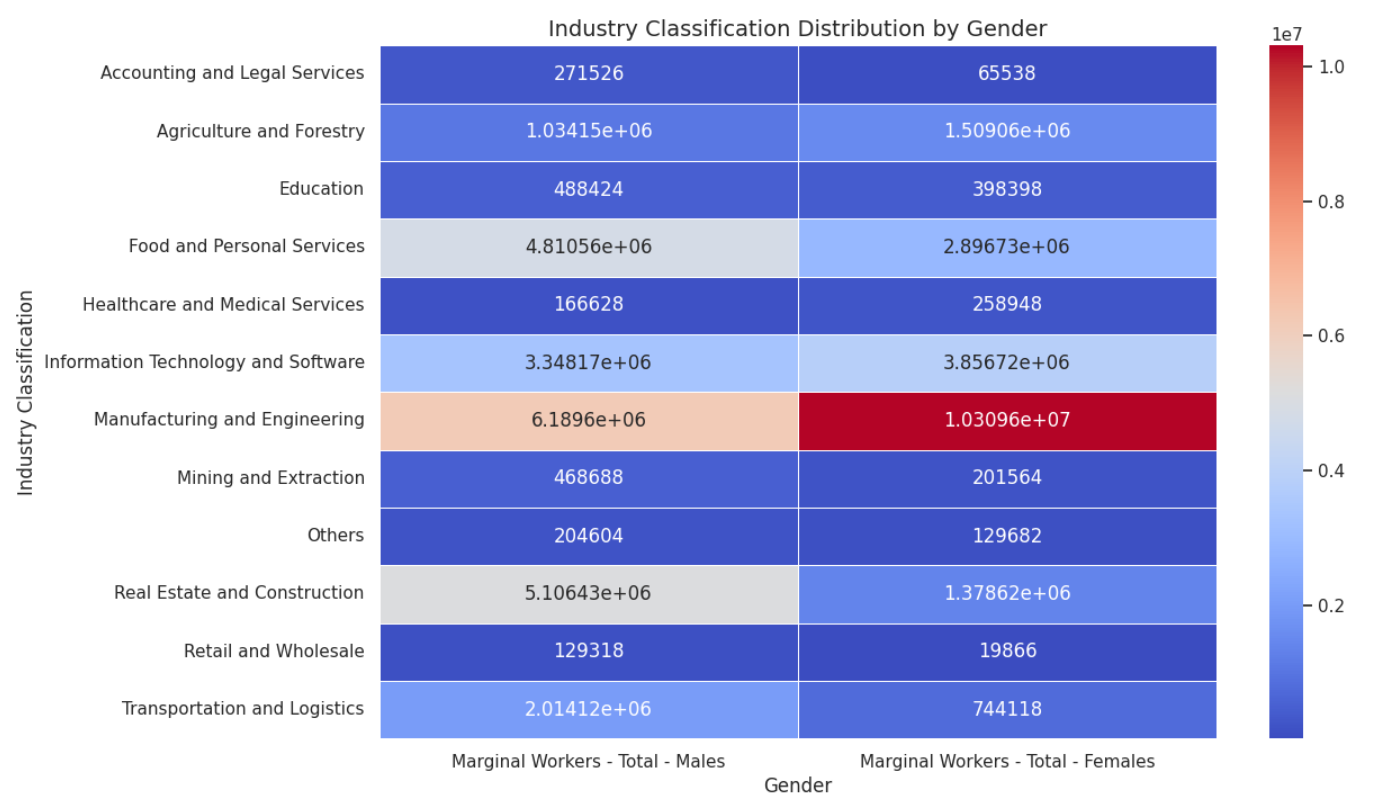
 **Consistent Gender Gap:**  
In every state, male main workers significantly outnumber females in both rural and urban segments.

 **High Employment in Large States:**  
States like Rajasthan, Gujarat, and Karnataka exhibit large-scale employment with substantial numbers in both rural and urban areas.

 **Rural Focus in Several States:**  
Many states (e.g., Assam, Bihar, Himachal Pradesh) show higher rural worker counts compared to their urban numbers, indicating a predominantly rural workforce.

**INDUSTRY CLASSIFICATION WITH MAIN AND MARGINAL WORKERS**





 **Food and Personal Services Dominate Main Workers:**

* This industry has the highest number of **Main Workers - Males (~48.34M)** and a significant number of **Females (~8.11M)**, making it one of the most gender-diverse industries.

 **Manufacturing and Engineering Leads in Marginal Workers:**

* In the **Marginal Workers** category, **Manufacturing and Engineering** stands out with **~6.19M Males** and **~10.3M Females**, indicating a higher female workforce in temporary or seasonal roles.

 **IT and Software Industry Shows More Stability:**

* The **Main Worker** count in **IT and Software (~21.76M Males, ~6.41M Females)** is significantly higher than **Marginal Workers (~3.34M Males, ~3.85M Females)**, indicating that this industry provides more stable, full-time jobs.

 **Agriculture and Forestry Has a High Female Marginal Workforce:**

* While the **Main Workers - Females** in **Agriculture & Forestry** are around **2.07M**, the **Marginal Workers - Females (~1.50M)** indicate a higher seasonal workforce, suggesting that many women engage in agriculture part-time.

COMPARISION OF DIVISION AND STATE WITH MAIN WORKERS

✅ **States like Rajasthan, Karnataka, and West Bengal have a strong Main Worker base, while Bihar and Odisha have a high number of Marginal Workers.**  
✅ **Urban areas have more Main Workers, while rural areas have a higher share of Marginal Workers.**  
✅ **Women are more likely to be Marginal Workers, particularly in rural areas.**  
✅ **Stable industries (manufacturing, services) employ more Main Workers, while agriculture and informal sectors have a larger Marginal workforce.**

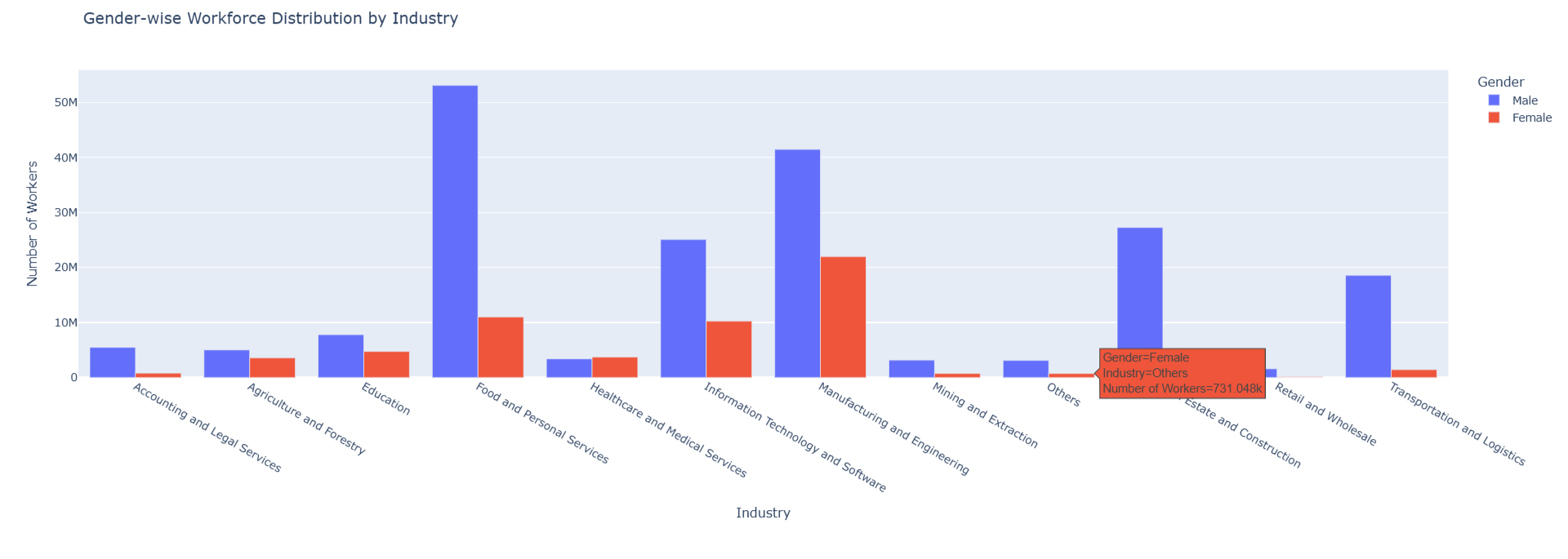
**GENDERWISE WORKFORCE DISTRIBUTION**

**Workforce Dominance** – Males dominate **Real Estate (27.2M), Transportation (18.5M), and Manufacturing (41.4M),** while females lead in **Manufacturing (21.9M), IT (10.2M), and Food Services (11M).**

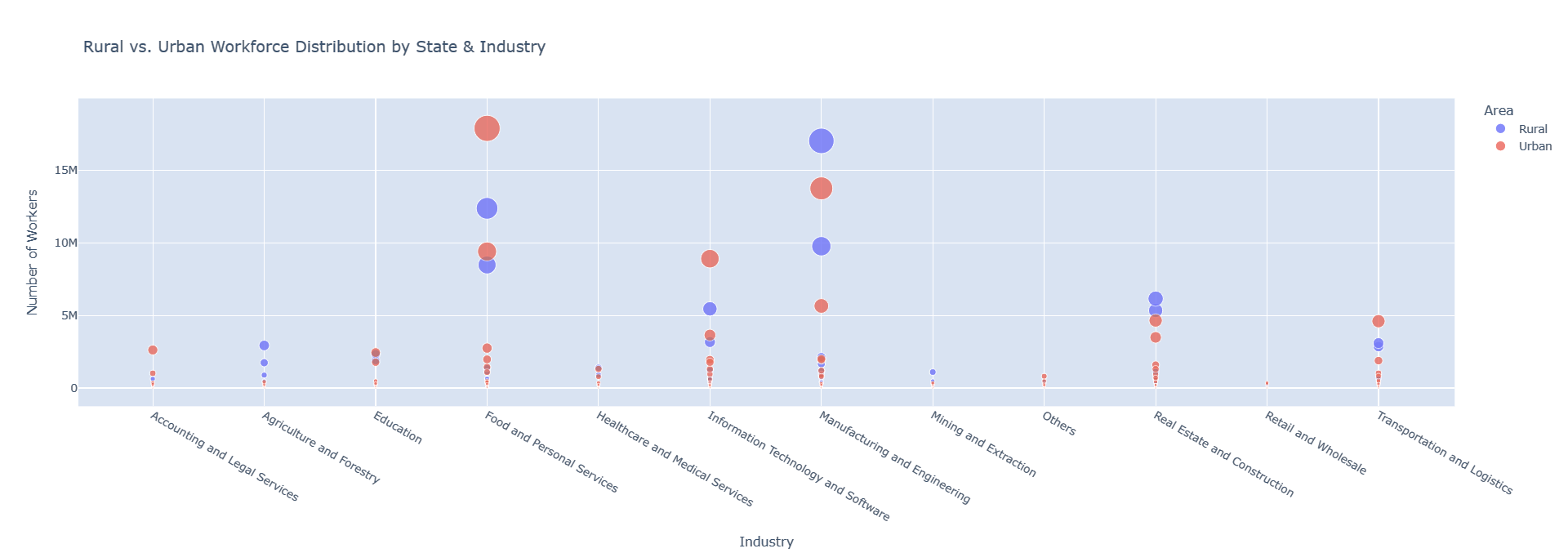
2️⃣ **High Female Participation** – IT (10.2M) and Healthcare (3.7M) show **growing female workforce,** while Manufacturing (21.9M) challenges traditional male dominance.

3️⃣ **Extreme Gender Gaps** – **Transportation (18.5M vs. 1.4M)** and **Real Estate (27.2M vs. 3.5M)** are highly male-dominated, while Retail has the **lowest female workforce (135K).**

4️⃣ **Strong Female Presence** – **Food & Personal Services (11M) and Education (4.7M)** see high female participation, with Agriculture also having significant female involvement.



**RURAL VS URBAN WORKFORCE**



1️⃣ **Rural-Dominated States** – **West Bengal (52.6M), Rajasthan (36.1M), and Odisha (4.9M)** have significantly higher rural workforce participation.

2️⃣ **Urban-Dominated States** – **Delhi (5.1M), Maharashtra (10.9M), and Tamil Nadu (9.3M)** have a strong urban workforce presence.

3️⃣ **Balanced Workforce States** – **Karnataka (5.8M rural vs. 6M urban) and Kerala (3.2M rural vs. 2.6M urban)** show a relatively equal distribution.

4️⃣ **Low Rural Workforce States** – **Goa (18K), Uttarakhand (155K), and Sikkim (173K)** have minimal rural workforce, likely due to urban-centric employment opportunities.

Tools and Technologies

Programming Language: Python

Libraries:

Data Processing: Pandas, NumPy

Machine Learning: Scikit-learn

NLP: NLTK / spaCy

Visualization: Plotly

Dashboard: Streamlit

Version Control: Git (with a public repository on GitHub)