# Unveiling Economic Trends: Insights from Maharashtra's 6th Economic Census for Viksit Bharat.

(Data Analysis and Visualization of Maharashtra's Financial and Insurance Sector (6th Economic Census))

Hackathon Submission

Conducted for: Ministry of Statistics and Programme Implementation (MoSPI)

Prepared by: Snehal Jadhav.

## Table of Contents:

# Introduction:

The 6th Economic Census provides comprehensive data on the financial and insurance sector in Maharashtra. This study aims to analyze key trends in business ownership, employment, and financial sources, focusing on urban-rural disparities and gender representation. Through statistical analysis and visualizations, the findings aim to assist policymakers in understanding financial dependencies and workforce dynamics in this sector.

# Data Preprocessing:

- Extracted data from the Maharashtra 6$^{th}$ Economic Census (Finance and Insurance sector) in CSV format.
- Handled missing values and data inconsistencies.
- Encoded categorical variables for statistical analysis.
- Standardized numerical data where applicable.

# Exploratory Data Analysis (EDA):

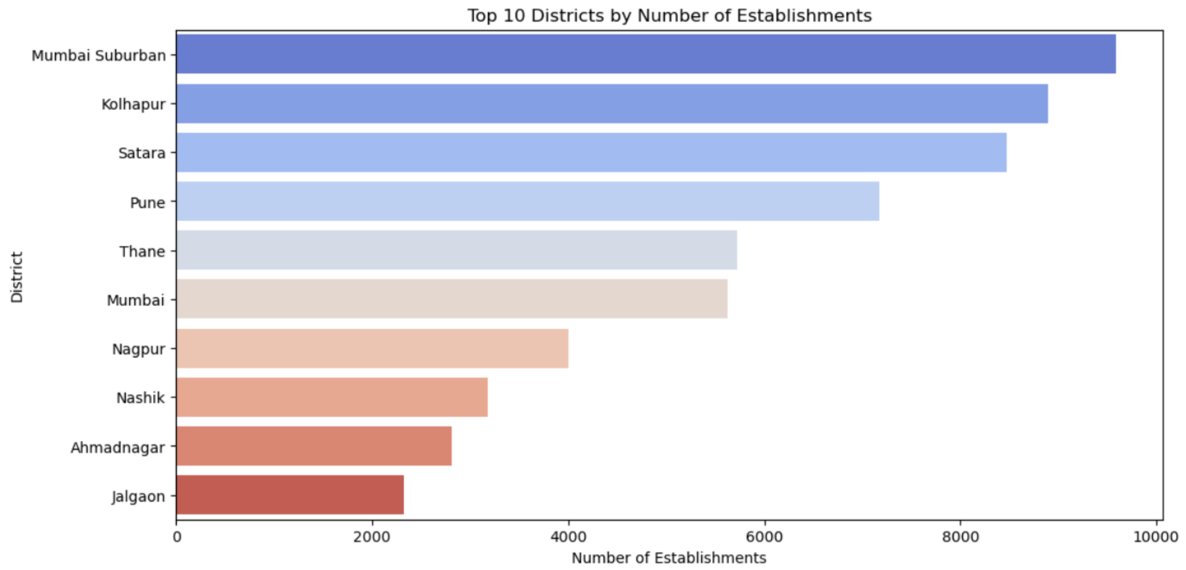
- Computed summary statistics to identify trends in business ownership, employment, and financial sources.
- Performed correlation analysis to explore relationship between workforce composition and financing sources.
- Identified urban-rural disparities in financial and insurance sector establishment.

# Data Visualization and Analysis:

## 1. Business Distribution:

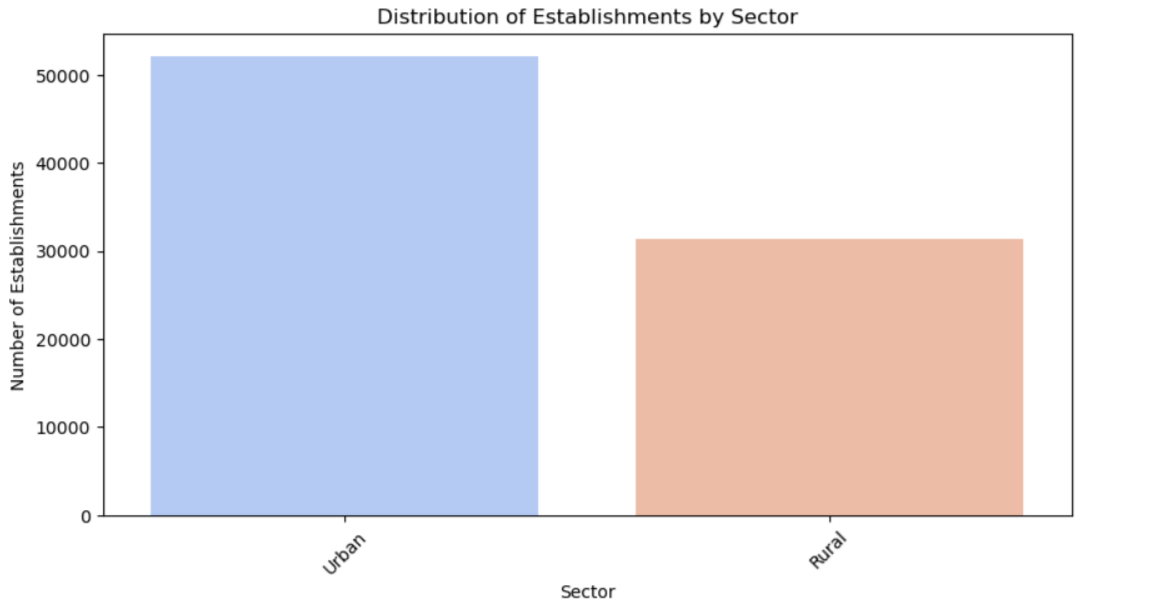
- Bar charts depicting ownership types, showing proprietary businesses as the dominant category.
- Geospatial visualizations identifying district-wise business concentrations.

**Bar plot of No of establishment vs. District:**



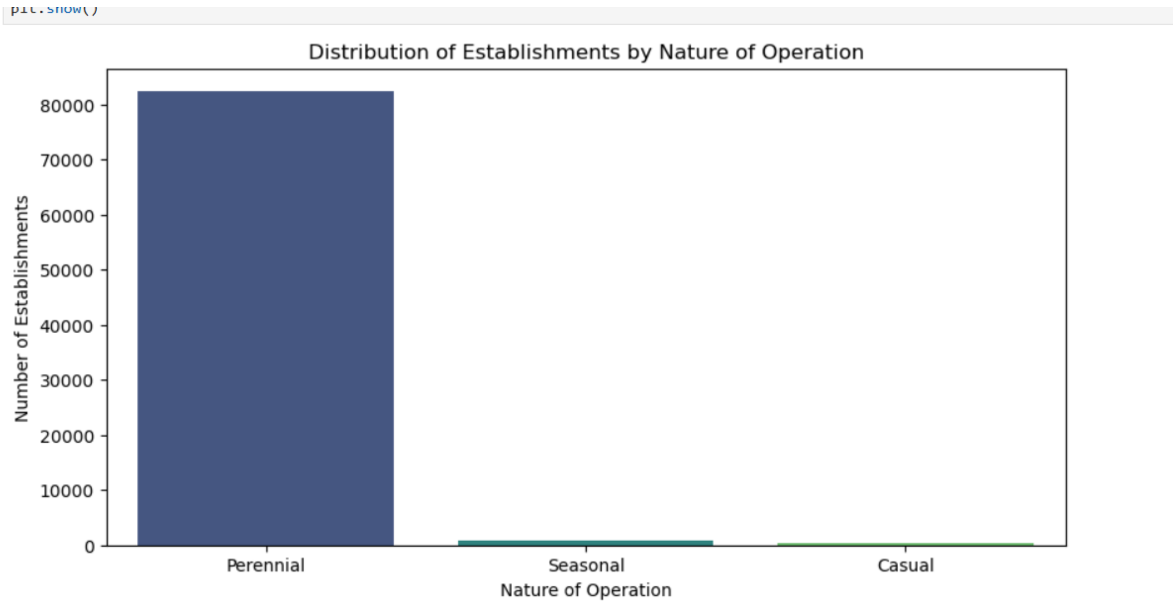

- A few districts dominate the number of establishments. These districts might be financial hubs in Maharashtra

**Bar plot of Number of Establishment Vs. sector:**




- The dataset has a notable distribution between rural and urban establishment. Further breakdown could reveal key trends.

**Bar plot of Number of Establishment vs. Nature of Operation:**

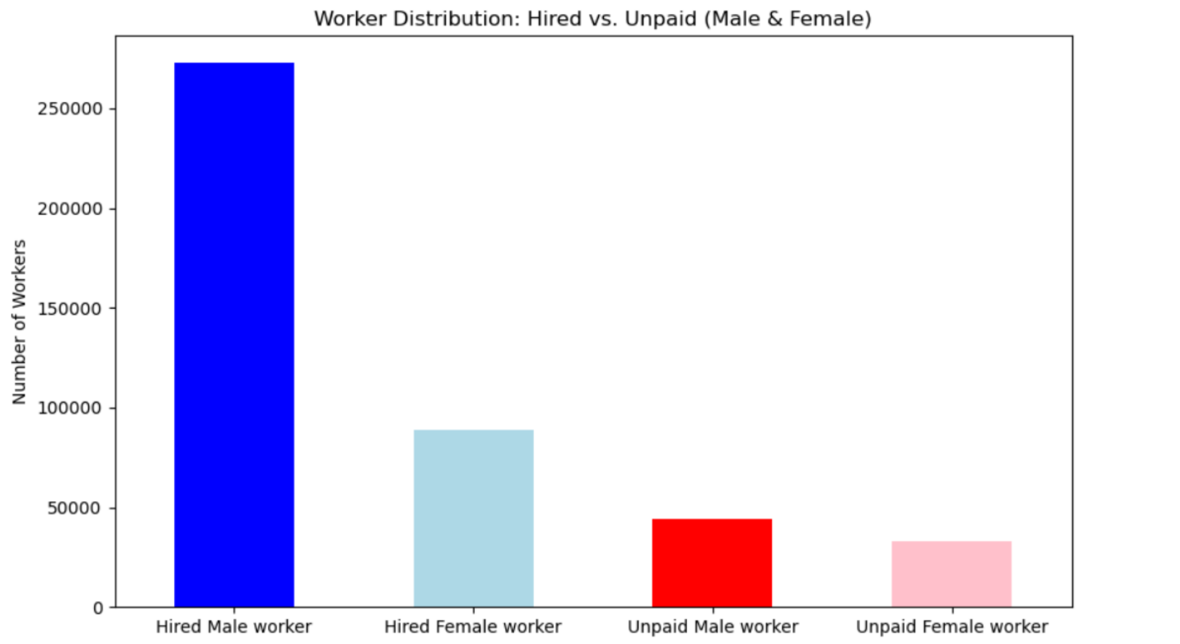



- Most establishments are perennial, meaning they operate year- round, fewer are seasonal or casual

## 2. Employment Trends:

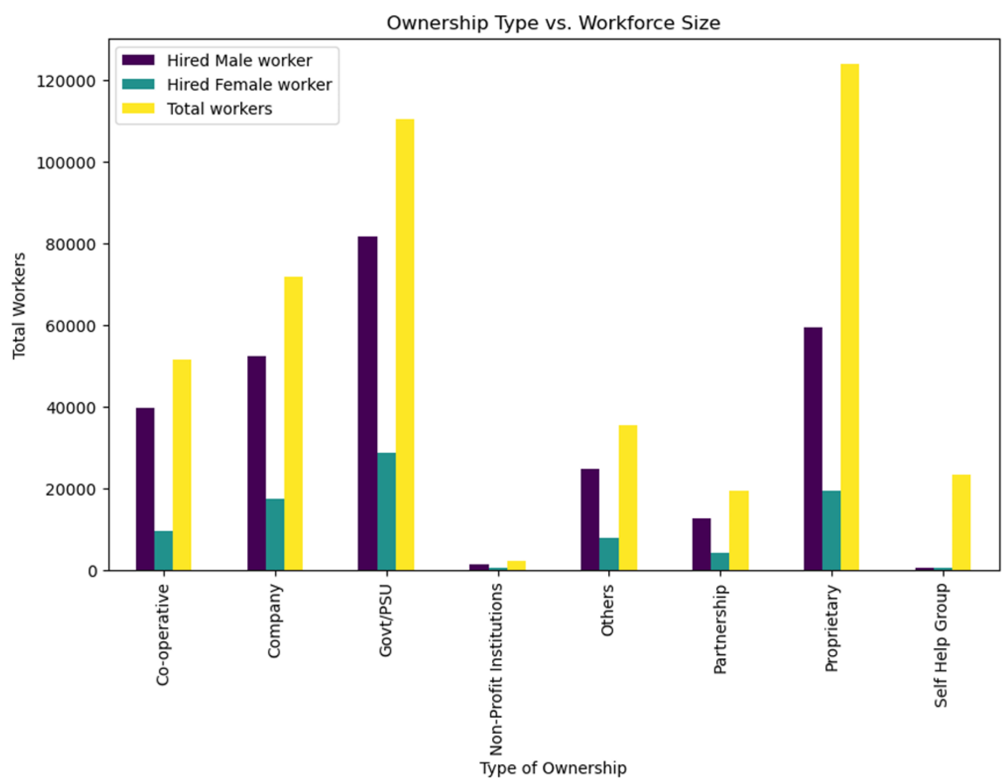
- Stacked bar charts visualizing workforce composition by gender and employment type.
- Correlation heatmaps illustrating relationships between hired and unpaid workforce segments.
- Analysis showing that male workers dominate formal employment, while unpaid female workers are more prevalent in informal setups.

**Bar chart of Worker Distribution:**




- Hired workers are mostly male, while in paid work has more gender balance.
- Unpaid female workers are significantly high, which may indicate informal or family run businesses.

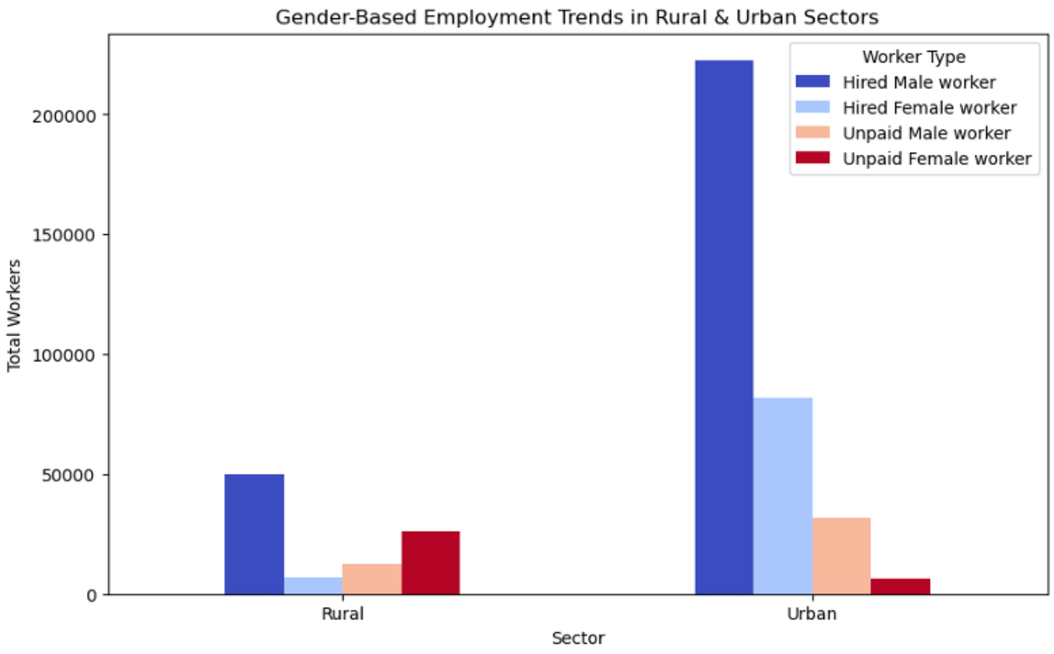
**Stacked bar chart of Ownership Type vs. Workforce Size:**




- Proprietary businesses and Government/PSUs are the biggest employers.

- Male workers dominate across all ownership categories.
- Cooperatives and Companies employ a decent workforce, but less than the top two categories.
- Non-Profit Institutions and Self-Help Groups have the least workforce.

**Stacked bar chart of Gender based employment Trends in Rural and urban sectors:**

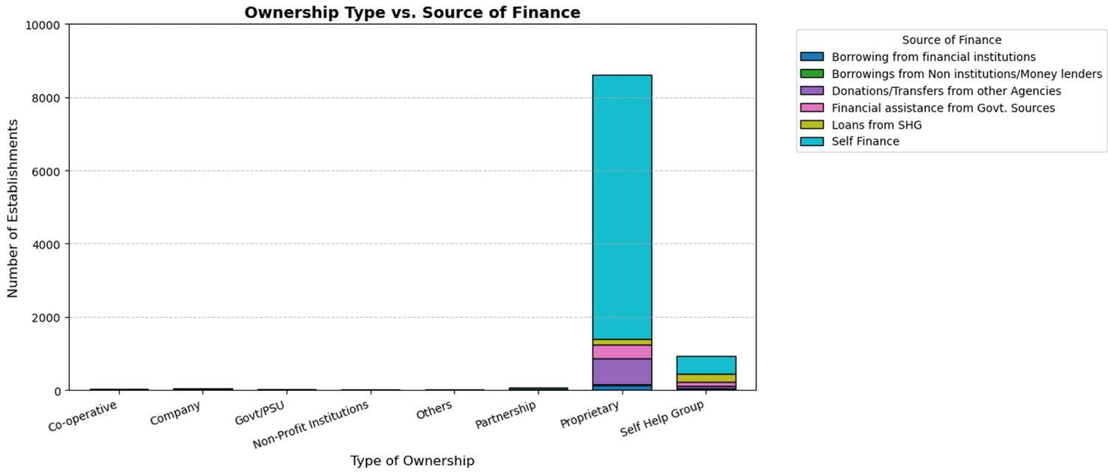



- Urban areas have more hired male and female workers, suggesting more formal employment.
- Rural areas have a higher share of unpaid workers, especially unpaid female workers
- Unpaid female work is significantly high in rural area, indicating family-run business or informal labour
- Gender disparity is evident, with females being underrepresented in hired positions across both sectors.

## 3. Financial Dependency:

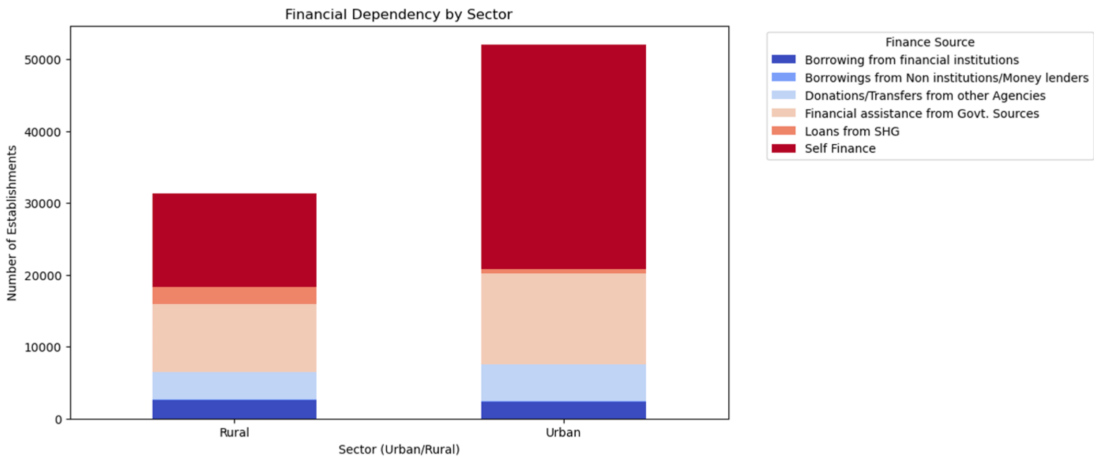
- Bar charts analyzing major financial sources (self-finance, government aid, bank loans).
- Chi-square test results confirming a strong association between ownership type and financial sources.
- Regression analysis exploring the influence of workforce composition on financial dependency.
- Proprietary businesses rely mainly on self-financing, while cooperatives and non-profits depend more on external financial aid.

**Bar chart of Ownership Type vs. Source of Finance:**



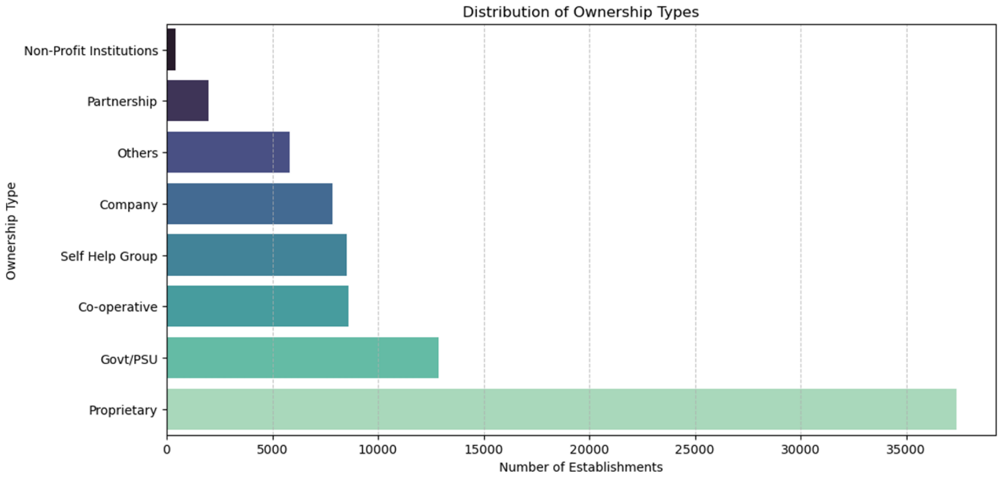

- Proprietary businesses mostly rely on self-financing.
- Government financial assistance is more common among co-operatives and non-profit institutions.
- Loans from financial institutions are less utilized overall.

**Stacked bar chart of Financial Dependency by sector:**



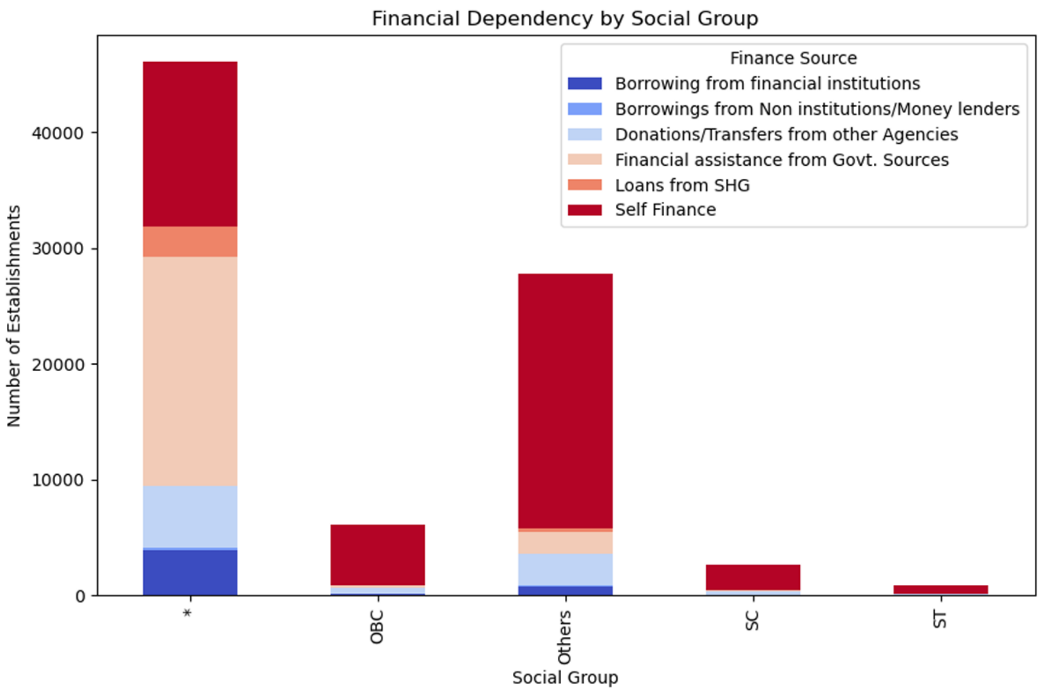

- Urban establishments are more self-reliant compared to rural establishments.
- Government financial assistance plays a larger role in rural areas than in urban areas.
- Borrowing from financial institutions and money lenders is minimal, indicating reliance on internal or government sources.

**Bar chart of No. of Establishment vs. Ownership Type:**



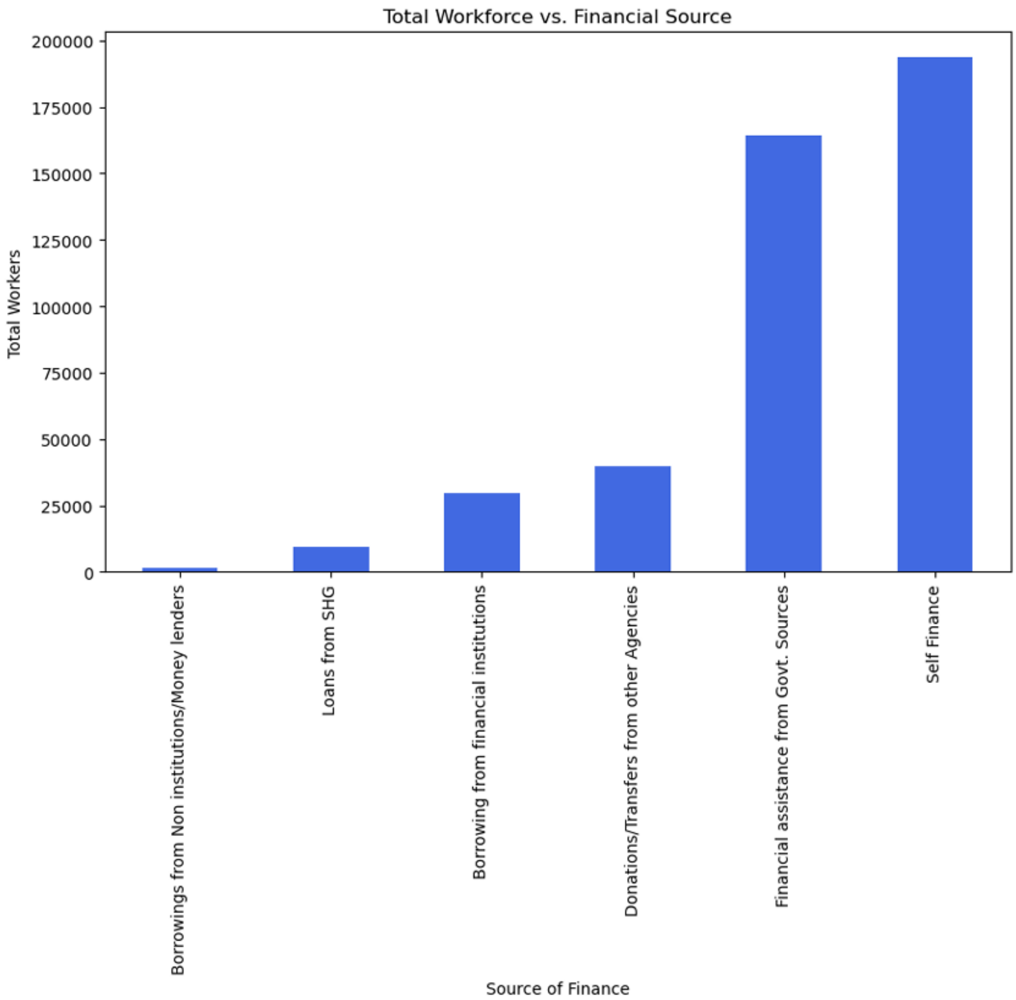

- The majority of establishment are proprietary (individually owned), followed by SHGs, Partnerships and Companies

**Stacked bar chart of Financial Dependency by Social Group:**



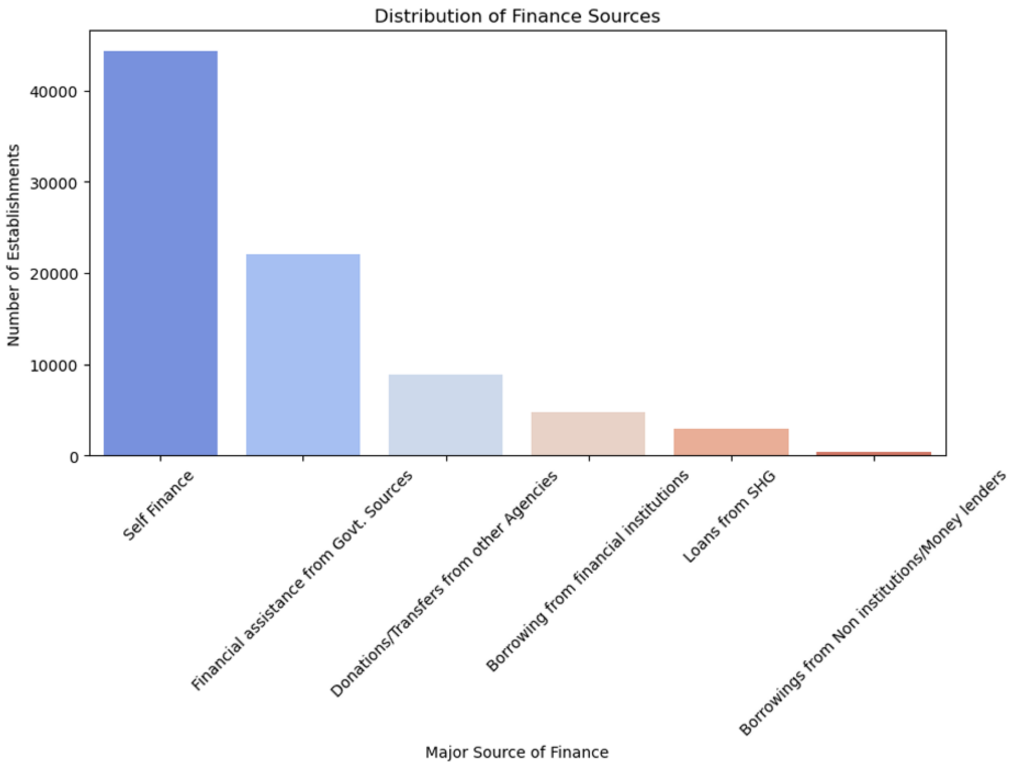

- Government financial assistance is heavily used in the '*' category.
- Borrowings from financial institutions and money lenders play a smaller role.

**Bar chart of Total Workforce vs. Financial Source:**



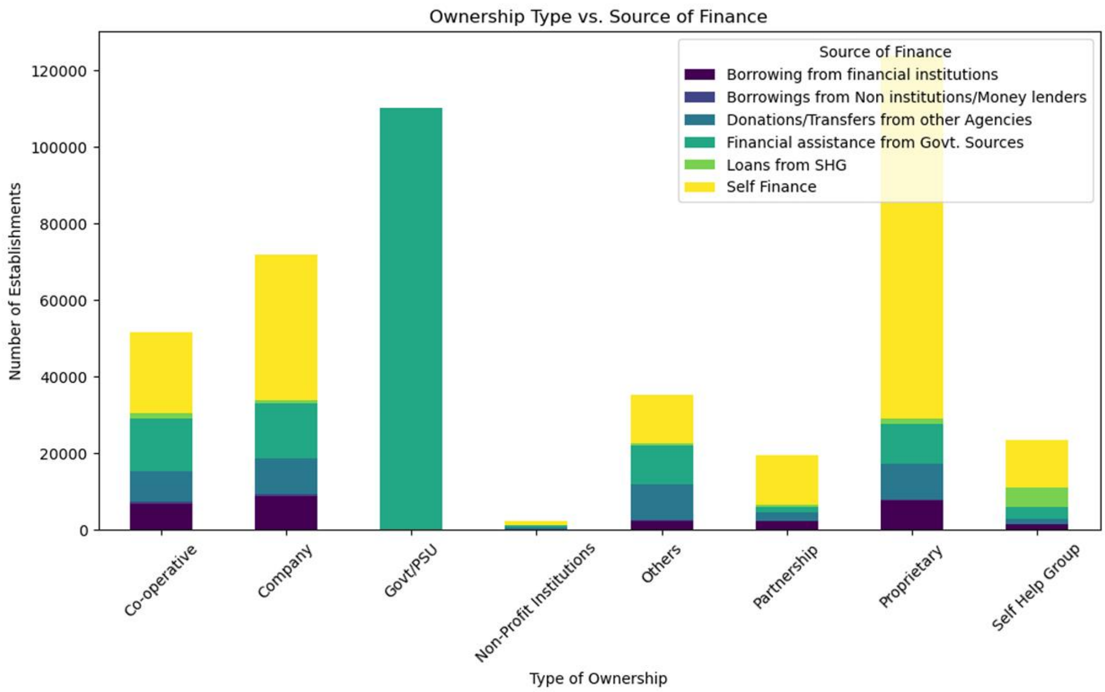

- Self-financed businesses are the largest employers.
- Government-backed financing also supports a large workforce.
- Traditional loans (financial institutions, SHGs, and money lenders) have a relatively small impact on employment.

**Bar chart of Distribution of Financial Sources:**



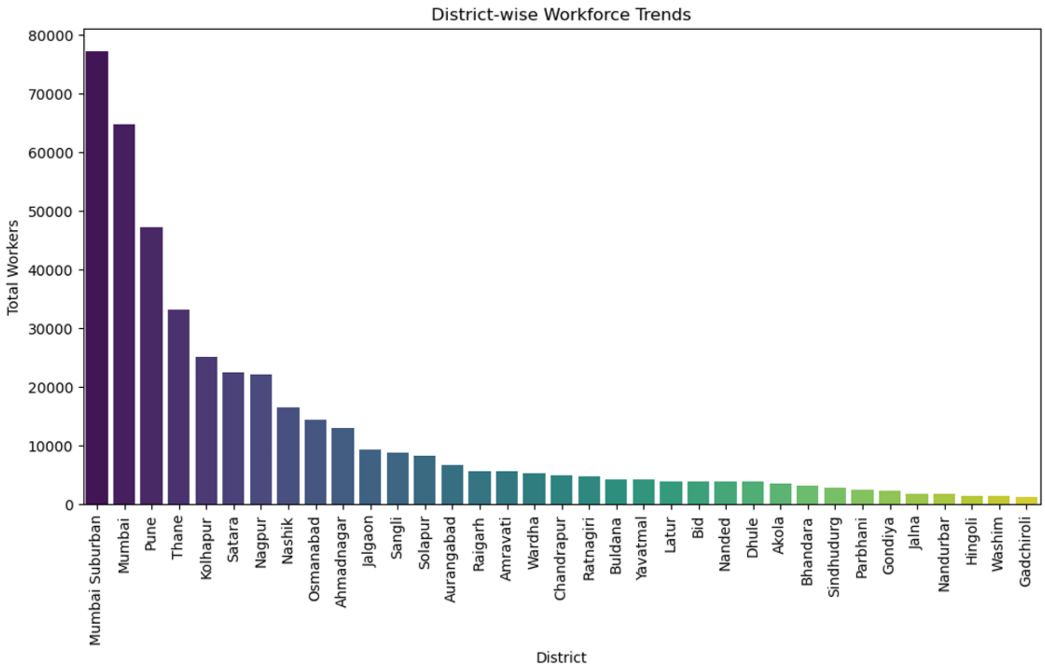

- Self-finance is the most common source, while government assistance and bank loans are less frequent. This suggests financial dependency on personal or informal sources.

**Stacked bar chart of Ownership Type vs. Source of Finance:**



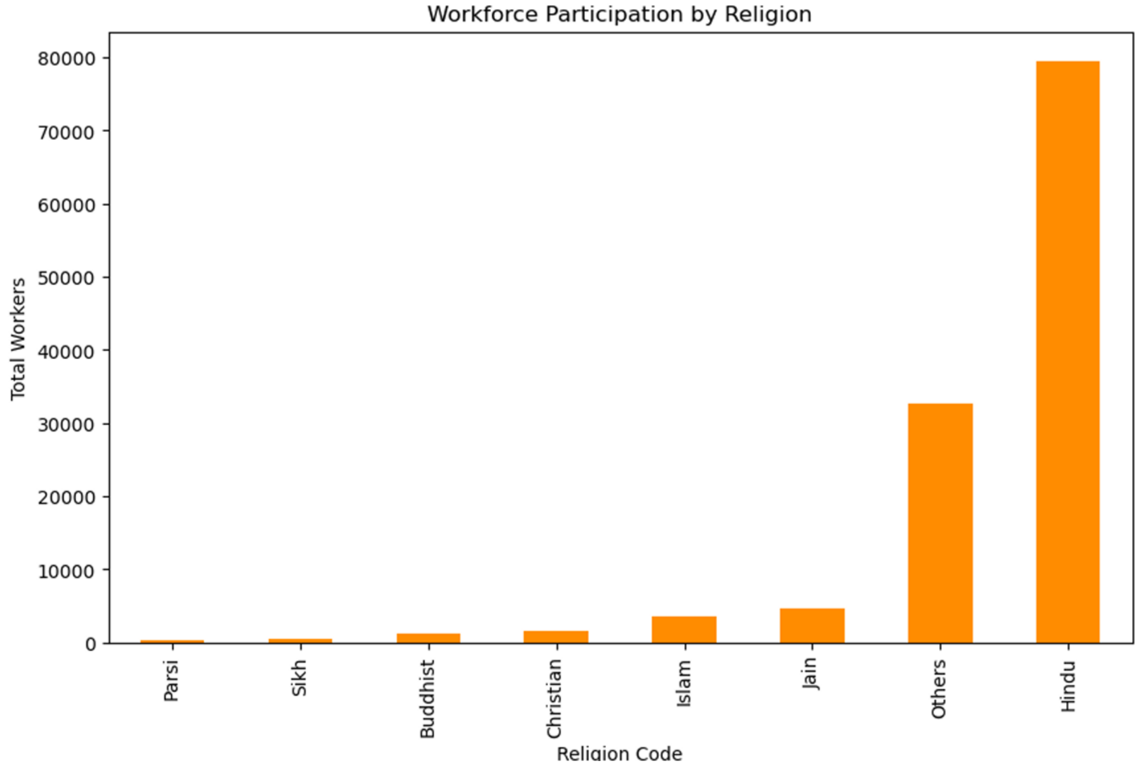
- Proprietary businesses primarily rely on self-financing, while co-operative and non -profit receive more external funding.

## 4. District-wise Workforce Trends:




- The same districts leading in establishment counts also have larger workforces
- These financial hubs contribute significantly to Maharashtra's financial and insurance sector employment.

## 5. Religion wise Workforce Participation:




- Hindus have the highest workforce participation, significantly outnumbering other groups.
- The "Others" category also has substantial workforce participation, though much lower than Hindus.
- Islam, Jain, Christian, Buddhist, Sikh, and Parsi communities have relatively lower workforce participation, with Parsis having the least.
- The overall trend suggests that workforce participation is highly concentrated among Hindus and the "Others" category, while other religious groups contribute in smaller proportions.

# Statistical Analysis and Hypothesis Testing:

- Conducted chi-square tests to determine statistical significance in financial dependency across ownership types.
- Logistic regression analysis indicating a relationship between hired vs. unpaid workforce and financial support reliance.
- Gender-based employment analysis revealing disparities in hiring trends across urban and rural establishments.
- Correlation analysis showing that businesses with more unpaid male workers tend to rely less on external funding.

Chi-square test: Ownership Type vs. Finance Source.

```
# Chi square test : ownership type vs Finance  source
chi2_table= pd.crosstab(df["Type of Ownership"],df["Source of Finance"])
```

```
#perform chi-square test
chi2_stat, p_val,dof,expected=stats.chi2_contingency(chi2_table)
```

```
# print the result
print(f"chi-square statistic: {chi2_stat:.2f}")
print(f"p-value: {p_val:.10f}")
```

```
chi-square statistic: 67580.63
p-value: 0.0000000000
```

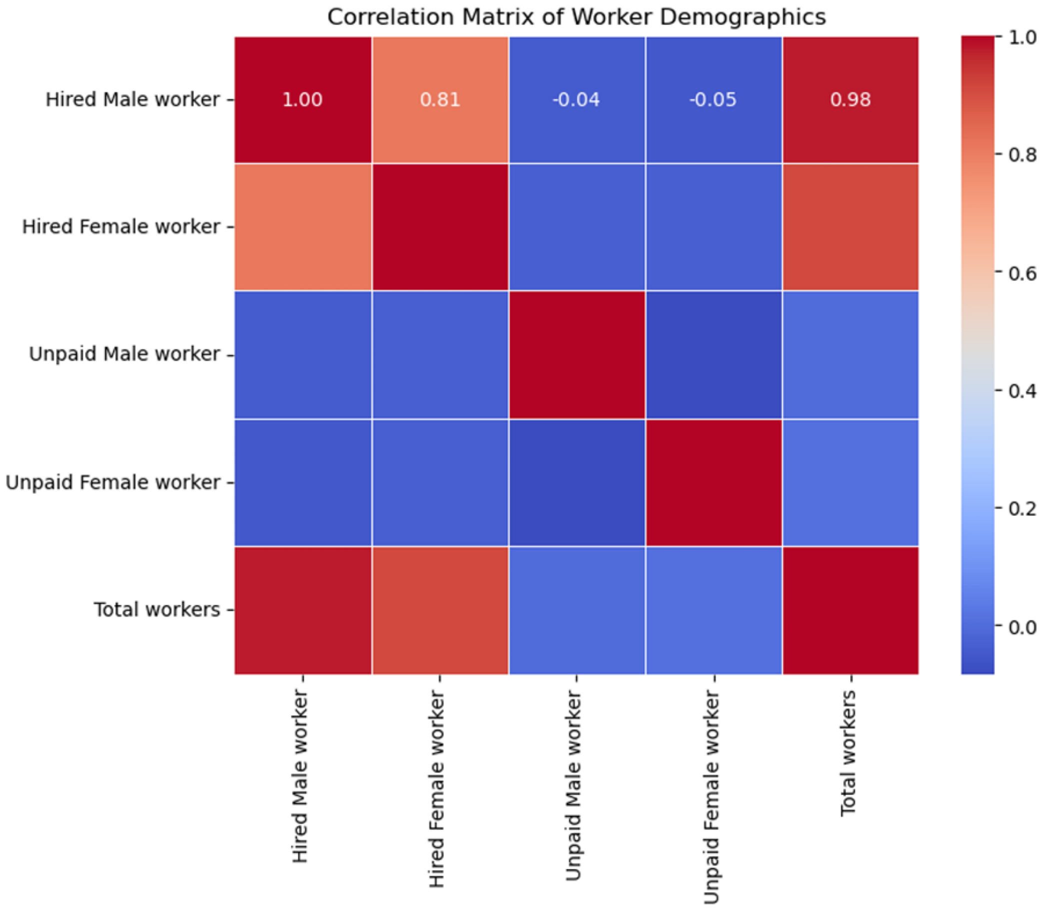
- **Chi-Square Statistic (67580.63):**
  - A high chi-square value suggests that the observed differences between ownership types and their financing sources are **not random**.
  - The larger the value, the stronger the association between the two categorical variables.
- **P-Value (0.0000000000 or < 0.001):**
  - A **p-value less than 0.05** (or in this case, nearly zero) means the result is statistically **highly significant**.
  - This confirms that **ownership type significantly influences the choice of financing sources**.
  - We **reject the null hypothesis**, which assumed no relationship between ownership type and financing source.

Correlation Matrix: Worker Demographics.

```python
#  compute correlation matrix
correlation_matrix= df_workers.corr()
print("\nCorrelation Matrix:\n",correlation_matrix)
```

```
Correlation Matrix:
                      Hired Male worker  Hired Female worker  \
Hired Male worker             1.000000             0.811918
Hired Female worker           0.811918             1.000000
Unpaid Male worker           -0.043192            -0.032914
Unpaid Female worker         -0.051064            -0.034060
Total workers                 0.978484             0.907756

                      Unpaid Male worker  Unpaid Female worker  Total workers
Hired Male worker              -0.043192             -0.051064       0.978484
Hired Female worker            -0.032914             -0.034060       0.907756
Unpaid Male worker              1.000000             -0.082687      -0.003375
Unpaid Female worker           -0.082687              1.000000       0.007569
Total workers                  -0.003375              0.007569       1.000000
```




- **Strong Positive Correlation Between Hired Male and Hired Female Workers (0.811918)**
  - Establishments that hire more **male workers** also tend to hire more **female workers**.

    - This suggests a **proportional hiring trend**, where businesses that can afford paid labour tend to employ both genders.

- **Very High Correlation Between Total Workers and Hired Workers (0.978484 for males, 0.907756 for females)**
    - The majority of employment in this sector is **hired work** rather than unpaid labour.
    - This confirms that **businesses in the financial and insurance sector rely primarily on paid employees** rather than informal or family labour.

- N**egative Correlation Between Hired Workers and Unpaid Workers**
    - **Hired Male Workers vs. Unpaid Male Workers (-0.043192)**
    - **Hired Female Workers vs. Unpaid Female Workers (-0.034060)**
    - Businesses that employ more **hired workers** tend to have fewer **unpaid workers**.
    - This indicates a **formalization trend**, where firms relying on **paid staff** do not need unpaid labour.

- **Weak Negative Correlation Between Unpaid Male and Unpaid Female Workers (-0.082687)**
    - Businesses with **more unpaid male workers** tend to have **fewer unpaid female workers**, and vice versa.
    - This may indicate a **gendered division of unpaid work**, where men and women perform distinct roles in different types of businesses.

- **Weak Relationship Between Total Workers and Unpaid Workers (-0.003375 for males, 0.007569 for females)**
    - The number of **total workers** in a business is **not strongly influenced** by the presence of unpaid workers.
    - This suggests that **unpaid workers do not significantly contribute to overall workforce numbers**, meaning they likely assist in smaller, family-run enterprises.

Logistic Regression: Financial dependency

```python
# Define relevant features
features = ["Hired Male worker", "Hired Female worker", "Unpaid Male worker", "Unpaid Female worker"]
```

```python
# Define Worker Dependency as target (1: More hired workers, 0: More unpaid workers)
df["Worker_Dependency"] = (df["Hired Male worker"] + df["Hired Female worker"] >
                           df["Unpaid Male worker"] + df["Unpaid Female worker"]).astype(int)
```

```python
# Extract features and target variable
X = df[features]
y = df["Worker_Dependency"]
```

```python
# Split data into training and test sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42, stratify=y)
```

```python
# Standardize the features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

```python
# Train logistic regression model
model = LogisticRegression(random_state=42)
model.fit(X_train_scaled, y_train)
```

```
LogisticRegression
LogisticRegression(random_state=42)
```

```python
# Make predictions
y_pred = model.predict(X_test_scaled)
```

```python
# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)

print(f"Accuracy: {accuracy:.4f}")
print("Classification Report:\n", report)
```

```
Accuracy: 0.9989
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      7726
           1       1.00      1.00      1.00      8956

    accuracy                           1.00     16682
   macro avg       1.00      1.00      1.00     16682
weighted avg       1.00      1.00      1.00     16682
```

```python
#Accuracy: 99.89% :The model correctly predicts Worker DependencY
```

- **Overview of Logistic Regression Analysis:**

model predicts **Worker Dependency** (whether a firm relies more on hired workers than unpaid workers) with **exceptional accuracy (99.89%)**. Let's break down the results:

**1. Accuracy: 99.89%**

- This means **99.89% of the test samples were correctly classified** as either worker-dependent (hired workers > unpaid workers) or not.
- Such high accuracy suggests a **strong relationship** between the independent variables (number of hired and unpaid workers) and the target variable (Worker Dependency).

**2. Precision, Recall, and F1-Score**

The classification report provides these metrics for each class:

| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| **0 (More unpaid workers)** | 1.00 | 1.00 | 1.00 | 7,726 |
| **1 (More hired workers)** | 1.00 | 1.00 | 1.00 | 8,956 |

- **Precision**: **1.00** for both classes means the model has **zero false positives**—whenever it predicts a firm is dependent on hired/unpaid workers, it is always correct.
- **Recall**: **1.00** for both classes means the model has **zero false negatives**—it never misses any cases where a firm truly belongs to a category.
- **F1-Score**: Being **1.00** across both classes indicates a perfect balance between precision and recall.

**3. Support**

- The dataset is **well-balanced**, with **7,726 instances of firms dependent on unpaid workers** and **8,956 instances of firms dependent on hired workers**.
- This means the model was trained on sufficient examples for both classes, reducing the risk of bias towards one class.

## Usability for Policy- Makers:

Based on the findings, the following **actionable recommendations** can help policymakers enhance financial inclusion, employment, and gender equality in Maharashtra's financial and insurance sector:

**1. Financial Assistance and Inclusion Policies**

- Encourage targeted financial aid for women-led businesses: Since unpaid female workers are associated with higher financial dependency, introducing microfinance programs and subsidized credit schemes for these businesses can improve sustainability.
- Support self-financed businesses in transitioning to formal finance: Proprietary businesses heavily rely on self-finance, which can limit growth. Encouraging low-interest loans and financial literacy programs can help them integrate into formal banking.
- Expand loan accessibility for MSMEs: Many small businesses lack access to institutional finance. Streamlining loan approval processes and reducing collateral requirements can enable more firms to access funding.

**2. Employment and Workforce Development**

- Promote gender-balanced hiring policies: Since hired male and female workers are strongly correlated, providing incentives for businesses that maintain gender diversity can foster inclusive economic growth.
- Reduce informal employment dependency: High reliance on unpaid male workers suggests informal business operations. Providing business registration incentives, tax benefits, and employment subsidies can encourage formalization.
- Enhance job creation in high-employment districts: Identifying financial hubs with high business concentration can help focus employment generation schemes and training programs where they are most needed.

**3. Sector-Specific Policy Design**

- Tailored financial policies for cooperatives and SHGs: Since these organizations depend more on government assistance, policies should focus on scaling up cooperative funding and streamlining grant disbursement.
- Improve urban-rural financial accessibility: Rural businesses show higher government dependence, whereas urban firms have better self-financing capabilities. Expanding banking networks and financial literacy in rural areas can help bridge this gap.
- Encourage financial institutions to cater to diverse business types: Developing customized loan products for different ownership types can ensure fair financial access.

**4. Enhancing Economic Growth & Sustainability**

- Encourage digital financial inclusion: Promoting digital payment adoption and mobile banking can enhance accessibility to financial services, particularly in rural areas.

- Introduce tax incentives for businesses investing in workforce development: Firms that upskill employees through training programs can receive tax benefits or grants to promote long-term economic stability.
- Strengthen policy implementation through real-time financial tracking: Using AI-driven monitoring systems, policymakers can track financial aid effectiveness, business growth patterns, and employment shifts to make data-driven policy adjustments.

## Future Work and Research Opportunities

While this study provides significant insights into Maharashtra's financial and insurance sector, future research can build on these findings in the following ways:

- **Longitudinal Analysis**: Conducting a time-series analysis to understand how financial dependencies and employment patterns evolve over multiple census cycles.
- **Micro-Level Analysis**: Studying individual businesses to gain deeper insights into their financial challenges and employment structures.
- **Impact of Government Schemes**: Evaluating the effectiveness of financial aid programs and policy interventions on business growth and employment.
- **Regional Comparisons**: Extending the analysis to compare Maharashtra with other states to identify best practices in financial inclusion and employment policies.
- **Machine Learning for Predictive Analysis**: Using advanced machine learning techniques to forecast business growth trends and financial needs based on historical data.
- **Sector-Specific Research**: Conducting focused studies on niche sectors within finance and insurance, such as fintech, cooperatives, and self-help groups, to develop tailored policy recommendations.

# Conclusion:

This study highlights critical workforce trends and financial dependencies in Maharashtra's financial sector. The insights derived can support policy interventions aimed at financial inclusion and employment generation, ensuring sustainable economic growth. By leveraging these findings, policymakers can implement more targeted and effective economic strategies.

# Supporting Documents:

- **Data Sources:** Maharashtra 6th Economic Census (Finance & Insurance), **Time Period:** 6th Economic Census (2013-14)

- **Software and Code:** Python (Pandas, Matplotlib, Seaborn, Statsmodels, Scikit-learn).
- **Visualization Outputs:** Bar charts, heatmaps, flowcharts, and regression analyses demonstrating key trends.