Unemployment and Skill Gap Analysis in India

A State-Level Analysis using PLFS and AISHE Data

Name: Ashish Vasala

Institution: MLR institute of technology

Internship : Data management and analysis division , NITI Aayog

1. Introduction

Unemployment among educated youth has emerged as a key policy concern in India. While access to higher education has expanded significantly over the past decade, corresponding employment outcomes have not always improved proportionately. This raises critical questions regarding the alignment between educational outcomes and labour market demand.

This project aims to analyse unemployment patterns across Indian States and Union Territories and examine whether disparities in employment outcomes can be explained by differences in higher education enrolment. By integrating labour market indicators with education supply data, the study seeks to identify potential skill gaps and regional mismatches.

---

2. Objectives of the Study

The key objectives of this analysis are:

- To analyse state-level unemployment patterns in India
- To examine labour force participation and employment absorption
- To assess higher education enrolment as a proxy for graduate supply
- To identify states exhibiting potential skill mismatches
- To derive policy-relevant insights for workforce planning

---

3. Data Sources

3.1 Periodic Labour Force Survey (PLFS)

Source: National Data and Analytics Platform (NDAP), MoSPI

The following indicators were extracted at the State/UT level using **Usual Status (Principal Status + Subsidiary Status)**:

- Unemployment Rate (UR)
- Labour Force Participation Rate (LFPR)
- Worker Population Ratio (WPR)

To ensure policy comparability, data were aggregated for:

- Gender: *Person*
- Area: *Total (Rural + Urban)*
- Standard working-age groups

3.2 All India Survey on Higher Education (AISHE)

Source: NDAP, Ministry of Education

AISHE data on **estimated enrolment at various education levels** was used as a proxy for graduate supply. Relevant post-secondary education levels were selected and aggregated at the state level.

---

4. Methodology

The analysis followed a structured and reproducible methodology:

1. **Data ingestion** from official government sources
2. **Data cleaning and preprocessing**, including:
    - Standardisation of state names
    - Conversion of enrolment values to numeric format
    - Handling of missing values

3. **Aggregation of PLFS indicators** to obtain one observation per state
4. **Aggregation of AISHE enrolment data** across education levels and gender
5. **Merging of PLFS and AISHE datasets** at the state level
6. **Validation checks** to ensure logical and statistical consistency
7. **Exploratory analysis and visualisation** to derive insights

---

5. Data Cleaning and Preprocessing

PLFS Data

PLFS datasets are multi-dimensional, containing disaggregations by age group, gender, and area. To obtain policy-standard state-level indicators:

- Aggregation was performed across relevant dimensions
- State-level averages were computed
- Logical consistency was verified (e.g., WPR ≤ LFPR)

AISHE Data

AISHE enrolment data was available in long format by education level and gender. Cleaning steps included:

- Selection of relevant post-secondary education levels
- Conversion of enrolment values to numeric format
- Treatment of missing enrolment values as zero, assuming non-reporting
- Aggregation across gender and education levels

This resulted in a single **Graduate Supply** indicator per state.

---

6. Validation and Consistency Checks

Post-merge validation included:

- Coverage checks across States and Union Territories
- Missing value assessment
- Range validation of labour market indicators
- Logical consistency checks (WPR ≤ LFPR for all states)

- Duplicate state detection

All validation checks passed, confirming the analytical reliability of the merged dataset.

---

7. Key Findings

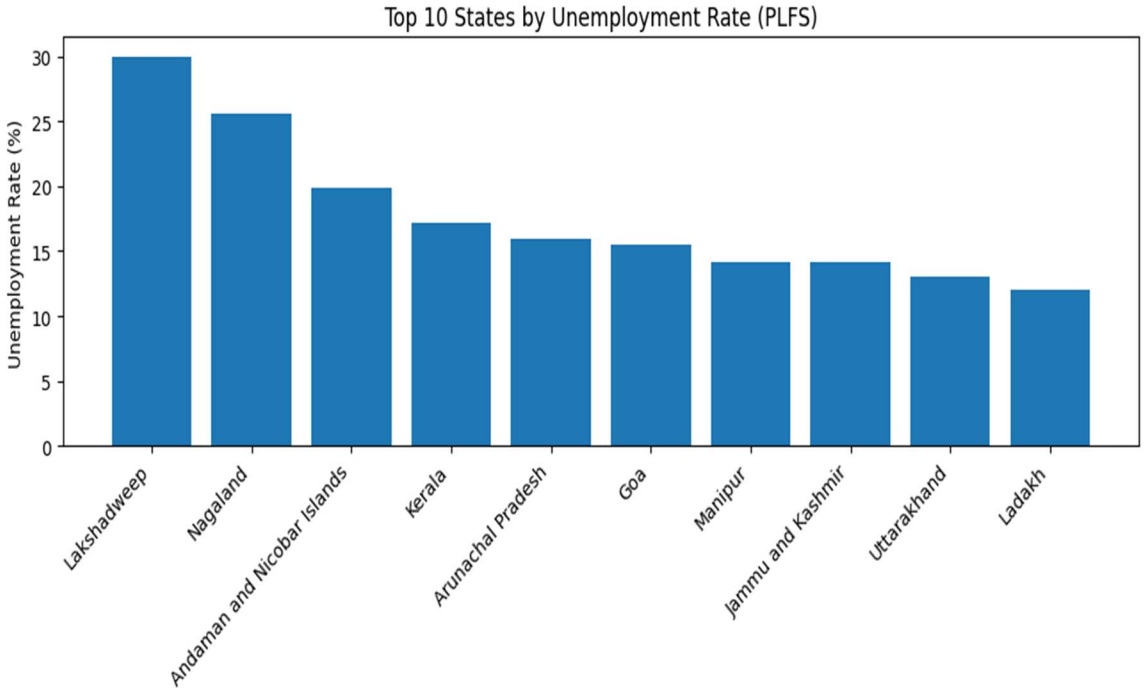

Figure 1: Top States by Unemployment Rate

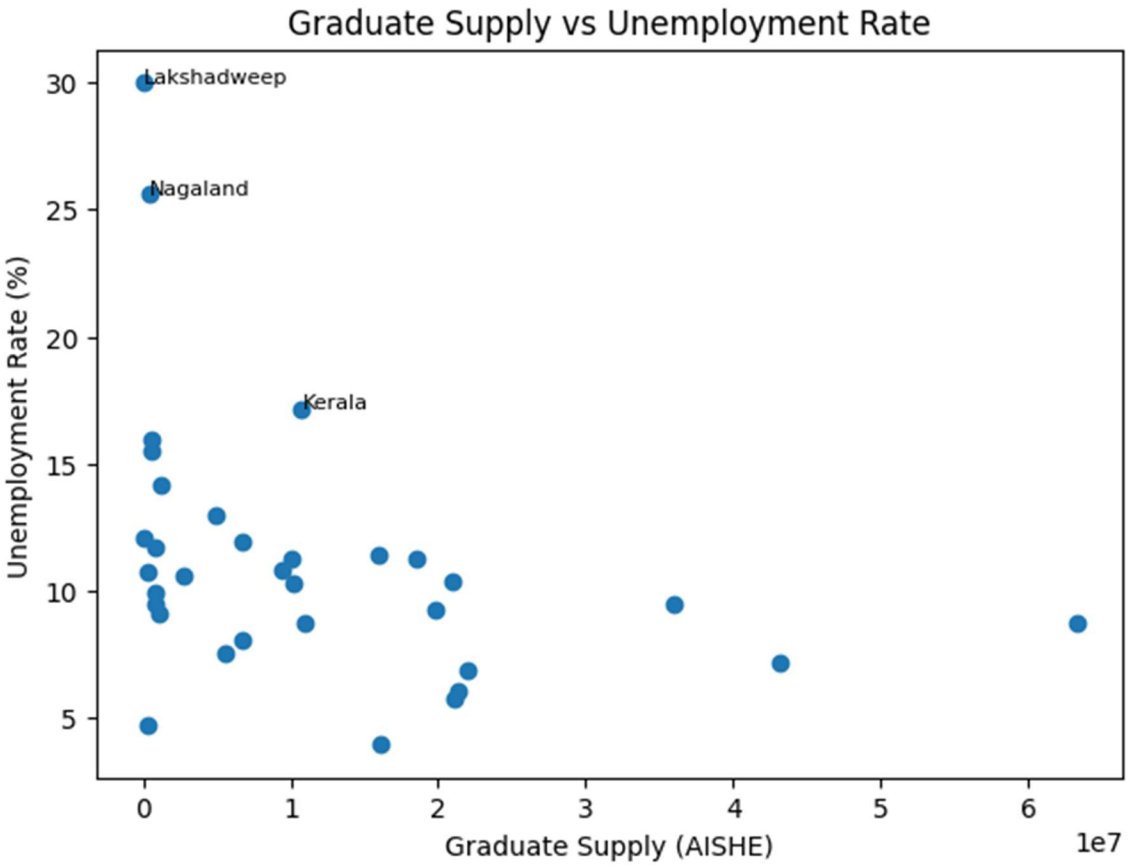

Figure 2: Graduate Supply vs Unemployment Rate

7.1 State-Level Unemployment Patterns

Significant variation exists in unemployment rates across states. Certain states exhibit unemployment rates substantially above the national average, indicating structural labour market challenges.

7.2 Graduate Supply and Employment Outcomes

The analysis reveals that states with higher graduate supply do not consistently exhibit lower unemployment rates. In several cases, high education output coexists with elevated unemployment, suggesting constraints in job absorption or skill relevance.

7.3 Labour Market Slack

The gap between Labour Force Participation Rate (LFPR) and Worker Population Ratio (WPR) highlights underutilised labour potential in certain states. Larger gaps indicate challenges in converting labour force participation into actual employment.

8. Skill Gap Insights

By jointly analysing PLFS and AISHE data, the study identifies:

- **Skill mismatch states**: High graduate supply coupled with high unemployment
- **Absorption-constrained regions**: High participation but low employment outcomes
- **Balanced states**: Relatively aligned education and employment outcomes

These patterns suggest that expanding education alone may not be sufficient without parallel improvements in job creation and skill alignment.

---

9. Policy Recommendations

Based on the findings, the following recommendations are proposed:

1. **Strengthen Industry–Academia Linkages**
   Curriculum design and training programs should be better aligned with regional labour market demand.
2. **Region-Specific Skill Interventions**
   Skill development and apprenticeship programs should be targeted at states exhibiting the largest education–employment mismatch.
3. **Data-Driven Workforce Planning**
   Integrated use of PLFS and AISHE data can support proactive planning of higher education capacity and skill development initiatives.

---

10. Limitations

- The analysis is conducted at the state level and does not capture intra-state or district-level heterogeneity.
- AISHE enrolment data is used as a proxy for graduate supply and does not directly measure employable skills.
- Missing enrolment values due to non-reporting were conservatively handled and may understate actual supply in some regions.

11. Conclusion

This study demonstrates the value of integrating labour market and education datasets to identify skill gaps and employment challenges. The findings highlight that higher education expansion must be complemented by targeted skill development and job creation strategies. The approach adopted in this project offers a scalable framework for evidence-based policy analysis.

Appendix A: Code Proofs

```
import pandas as pd

path = "/content/drive/MyDrive/PLFS_Project/data_raw/"

ur = pd.read_csv(path + "ur_state.csv")

lfpr = pd.read_csv(path + "lfpr_state.csv")

wpr = pd.read_csv(path + "wpr_state.csv")

edu = pd.read_csv(path + "education_unemployment.csv")
```

```
import matplotlib.pyplot as plt

df_sorted = df.sort_values('UR', ascending=False).head(10)

plt.figure(figsize=(10,5))

plt.bar(df_sorted['State'], df_sorted['UR'])

plt.xticks(rotation=45, ha='right')

plt.ylabel('Unemployment Rate (%)')

plt.title('Top 10 States by Unemployment Rate (PLFS)')

plt.tight_layout()

plt.show()
```

```
final_df.shape
final_df.isna().sum()
final_df[final_df['Graduate_Supply'].isna()][['State']]
final_df[['UR', 'LFPR', 'WPR', 'Graduate_Supply']].describe()
(final_df['WPR'] <= final_df['LFPR']).all()
final_df['State'].duplicated().sum()
#plot checking
import matplotlib.pyplot as plt

plt.figure(figsize=(7,5))
plt.scatter(final_df['Graduate_Supply'], final_df['UR'])

for _, row in final_df.iterrows():
   if row['UR'] > final_df['UR'].quantile(0.9):
      plt.text(row['Graduate_Supply'], row['UR'], row['State'], fontsize=8)

plt.xlabel('Graduate Supply (AISHE)')
plt.ylabel('Unemployment Rate (%)')
plt.title('Graduate Supply vs Unemployment Rate')
plt.show()
#correlation
final_df[['Graduate_Supply', 'UR']].corr()
assert not final_df.empty, "Final dataset is empty!"
```