****

**Goa Institute of Management**

**Report on**

**Child Marriage Vulnerability in Chhattisgarh:**

**A Data-Driven Approach**

**Group no. - 21**

**Rhea Pancholi – B20240105**

**Moksh Grover – B2024096**

**Soham Pal – B2024117**

**Darshan Pradeep – B2024078**

**Saanya Lasod – B2024106**

**Chetan Ahuja – B2024077**

**Acknowledgement**

We would like to express our heartfelt thanks to everyone who has contributed to the successful completion of this project. First and foremost, we extend our deepest gratitude to our project guide and mentor **Prof. Sumit Tripathi**, for providing us with invaluable guidance, feedback, and support throughout the project. His knowledge and experience were instrumental in helping us navigate the complexities of the project, and we truly appreciate his dedication to our success. We would also like to acknowledge the contributions of **Prof. P. Balasubramanyam** for the valuable support throughout the Give Goa project tenure. We would also like to thank the team at **CRY organisation**, our client for the project, for their guidance and suggestions throughout the journey. They were very clear with their requirements, which enabled us to move in the right direction. Their contribution is immense for the success of the project.

Contents

[**Abstract: 6**](#_Toc192103598)

[**Introduction 7**](#_Toc192103599)

[**Literature Review 9**](#_Toc192103600)

[**Prevalence and Trends 9**](#_Toc192103601)

[**Determinants of Child Marriage 9**](#_Toc192103602)

[**Government and Non-Governmental Interventions 9**](#_Toc192103603)

[**Data and Methodology: 11**](#_Toc192103604)

[**Data Collection & Sources: 11**](#_Toc192103605)

[**Weight based scoring model: 11**](#_Toc192103607)

[**Risk Categorization Model 12**](#_Toc192103608)

[**Data Processing & Visualization 12**](#_Toc192103609)

[**Model Validation & Cross-Verification 13**](#_Toc192103610)

[**Clustering Approach 13**](#_Toc192103611)

[**Feature Selection and Transformation 13**](#_Toc192103612)

[**Clustering Approach 13**](#_Toc192103613)

[**Cluster Analysis 15**](#_Toc192103614)

[**Machine Learning Models 15**](#_Toc192103615)

[**Model Evaluation and Validation 16**](#_Toc192103616)

[**Limitations of the Study 18**](#_Toc192103617)

[**Research Findings And Insights: 20**](#_Toc192103618)

[**Findings Based on Weighted Score Approach 20**](#_Toc192103619)

[**Analysis of Cluster Vulnerability for Child Marriage** 21](#_Toc192103620)

[**Analysis of Principal Component Analysis (PCA) Feature Importance** 21](#_Toc192103621)

[**Analysis of Clusters:** 23](#_Toc192103622)

[**Analysis of Key Features by Clusters:** 23](#_Toc192103623)

[**Key Risk Factors and Defined Thresholds** 28](#_Toc192103629)

[**Cluster Analysis with Quantitative Rankings** 29](#_Toc192103630)

[**Summary of Key Metrics and Rankings** 30](#_Toc192103631)

[**Analysis of Vulnerability by Key Factors and their implication:** 33](#_Toc192103634)

[**Analysis of Random Forest and XGBoost Models in Predicting Child Marriage Vulnerability** 40](#_Toc192103636)

[Conclusion & Recommendation: 43](#_Toc192103637)

**Abstract:**

This study examines child marriage vulnerability in Chattisgarh through a data-driven lens, integrating statistical modeling, clustering techniques, and machine learning algorithms. Utilizing a dataset of 4,644 records from partner organizations, field surveys, and government sources, the research identifies key socioeconomic, educational, and demographic factors contributing to child marriage risk. A weighted score approach quantifies individual vulnerability, while Principal Component Analysis (PCA) highlights the most influential predictors. K-Means clustering segments individuals into five risk categories — Extreme, High, Moderate, Mild, and Low Vulnerability — offering a nuanced understanding of regional disparities. Furthermore, machine learning models, including Random Forest and XGBoost, achieve 99.08% accuracy in predicting child marriage vulnerability, with Anganwadi enrollment, Early Childhood Education (ECE), and parental occupation stability emerging as the strongest predictors. The study underscores the urgent need for targeted, region-specific interventions, advocating for a multi-sectoral approach that combines educational support, economic empowerment, and community-based advocacy. These findings contribute to the growing body of knowledge on child marriage in India, providing actionable insights for policymakers and stakeholders to design data-informed strategies that protect at-risk populations and promote child welfare.

# **Introduction**

Child marriage remains a deeply entrenched social issue in India, perpetuated by a complex interplay of socioeconomic, cultural, and gender-based factors. Despite the enactment of the Prohibition of Child Marriage Act (2006) and numerous government and non-governmental interventions, the practice persists, particularly in rural and marginalized communities. According to UNICEF (2021), India accounts for one-third of the world’s child brides, with 223 million child brides, highlighting the scale of this challenge. Although the national child marriage rate has decreased from 47% to 27% over the past decade, regional disparities remain stark, with states like Bihar and Rajasthan exhibiting persistently high rates.

This study focuses on assessing child marriage vulnerability in Chattisgarh using a data-driven approach, integrating quantitative scoring models, clustering techniques, and machine learning algorithms. A dataset of 4,644 records, sourced from partner organizations, field surveys, and government records, underpins this research, encompassing demographic, socioeconomic, educational, and regional variables. The study employs a weighted score approach to quantify individual vulnerability levels, with indicators such as gender, school attendance, parental education, and household economic status assigned specific weights based on empirical research and expert validation.

Furthermore, Principal Component Analysis (PCA) was utilized to reduce dimensionality and identify the most influential features contributing to child marriage risk. Clustering methods, particularly K-Means, were adopted to categorize individuals into distinct risk groups — Extreme Vulnerability, High Vulnerability, Moderate Vulnerability, Mild Vulnerability, and Low Vulnerability. The choice of K-Means over Agglomerative Clustering and DBSCAN was justified by its superior segmentation clarity and interpretability, as evidenced by silhouette scores and within-cluster variance.

In addition to clustering, machine learning models — Random Forest and XGBoost — were employed to predict child marriage vulnerability with remarkable accuracy rates of 99.08%. These models revealed that lack of Anganwadi enrollment, limited access to Early Childhood Education (ECE), and parental occupation stability were the strongest predictors of child marriage risk. The findings underscore the urgent need for targeted interventions, combining educational support, economic empowerment, and community-based advocacy to combat this multifaceted issue.

By adopting a rigorous, data-centric methodology, this research aims to bridge the gap between statistical insights and actionable policy recommendations. It advocates for a multi-sectoral approach, emphasizing region-specific strategies to address the root causes of child marriage and protect vulnerable populations. Ultimately, this study contributes to the growing body of knowledge on child marriage in India, offering empirical evidence to inform more effective, evidence-based interventions.

# Literature Review

Child marriage remains a pressing social issue in India, despite legal prohibitions and ongoing intervention efforts. This literature review synthesizes insights from four key sources: the Times of India article (2014), the NCPCR report (2023-24), the ICRW and UNICEF district-level study (2016), and UNICEF’s Child Marriage Country Profile (2021). Together, these works provide a comprehensive understanding of the prevalence, causes, and countermeasures associated with child marriage in India.

## Prevalence and Trends

The Times of India (2014) highlights an unexpected trend in Chhattisgarh, where child marriage among males outpaces that of females, with 10.4% of males marrying before the legal age of 21 compared to 4.7% of females under 18. This pattern is most pronounced in tribal regions, where cultural traditions perpetuate early marriage.

The UNICEF (2021) report adds broader context, revealing that India accounts for one-third of the world’s child brides, with 223 million child brides. Although the national child marriage rate has declined from 47% to 27% over the past decade, the report underscores how the COVID-19 pandemic exacerbated the issue by driving families into economic distress, which, in turn, increased the risk of child marriage.

The NCPCR (2023-24) report provides granular, district-level data, emphasizing states like Bihar and Rajasthan with persistently high child marriage rates. The report outlines government strategies, including virtual review meetings and community awareness initiatives, aimed at curbing these trends.

## Determinants of Child Marriage

The ICRW and UNICEF (2016) study delves into the socioeconomic determinants of child marriage, identifying poverty, lack of education, and gender norms as primary drivers. The study highlights how districts with lower literacy rates and limited economic opportunities show higher incidences of child marriage. Moreover, social pressure and deeply rooted patriarchal values further entrench this practice.

The UNICEF (2021) profile reiterates these findings, adding that rural-urban disparities play a significant role. Rural areas witness a higher prevalence of child marriage due to limited access to education and resources. The pandemic has widened this gap, pushing vulnerable families to resort to early marriage as a coping mechanism.

## 

## Government and Non-Governmental Interventions

The NCPCR (2023-24) report showcases various government efforts, such as the establishment of child protection committees, community engagement programs, and the enforcement of the Prohibition of Child Marriage Act (2006). The report emphasizes the need for a multi-sectoral approach, involving education, healthcare, and social welfare systems.

The ICRW and UNICEF (2016) study highlights non-governmental interventions, including life-skills training for girls, community-based advocacy, and incentive programs encouraging families to keep their daughters in school. Such grassroots efforts have shown promise in reducing child marriage rates in high-risk districts.

Collectively, these sources underscore the complexity of child marriage in India, revealing a nexus of socioeconomic, cultural, and gender-based factors. While national trends show progress, regional disparities and emerging challenges, such as those posed by the COVID-19 pandemic, call for sustained, collaborative efforts between government bodies, NGOs, and local communities. Future research should focus on longitudinal data to evaluate the effectiveness of intervention programs and address evolving risks linked to economic instability and public health crises.

# 

**3.** **Data and Methodology:**

This study employs a structured, data-driven approach to assess child marriage vulnerability in Chattisgarh, integrating quantitative scoring, socio-economic analysis, and visualization.

## **Data Collection & Sources:**

## A dataset of 4,644 records was compiled from partner organizations, field surveys, and government records, covering demographic details (age, gender, caste, religion), household socio-economic status (income, occupation, housing conditions), educational status (school attendance, parental education), family structure, and regional factors.

**3.1 Weighted Score Approach**

### **Weight based scoring model:**

A weight-based scoring model was implemented to determine the vulnerability of each child based on multiple risk factors. The final vulnerability score for each child was computed as:

where:

* represents the value of the indicator for the individual,
* is the assigned weight for the indicator, and
* is the total number of indicators.

The weights assigned to each indicator were determined based on expert opinions, historical trends, and prior research studies. The specific weight distribution is provided in **Table 1**.

#### **Table 1: Weighting Criteria & Justification**

|  |  |  |
| --- | --- | --- |
| **Indicator** | **Weight(%)** | **Justification** |
| Gender | 15% | Girls face higher risk due to social norms. |
| Age | 10% | Younger children are more vulnerable. |
| School Attendance | 20% | Dropouts indicate increased risk. |
| Parental Education | 10% | Lower parental education correlates with traditional beliefs. |
| Parental Occupation | 10% | Informal-sector employment increases risk. |
| Household Economic Status | 10% | Poverty increases susceptibility to early marriage. |
| Single Parent / Orphan Status | 15% | Orphans/single-parent children face higher social vulnerability. |
| Caste & Religion Influence | 5% | Cultural traditions affect child marriage prevalence. |
| Housing Condition | 5% | Poor housing conditions correlate with financial instability. |

### **Risk Categorization Model**

Based on the final weighted scores, individuals were classified into three risk categories:

where is the assigned risk category for individual . This classification enables targeted interventions and policy measures for the most vulnerable populations.

### **Data Processing & Visualization**

Data visualization techniques were implemented using **Power BI** to generate insights. The following representations were utilized:

* **District-Wise Heatmaps**: Identified high-risk regions for focused interventions.
* **Bar Charts & Distribution Plots**: Analyzed gender-based and economic risk factors.
* **Scatter Plots & Correlation Analysis**: Evaluated relationships between socio-economic indicators and vulnerability scores.

### **Model Validation & Cross-Verification**

The model was validated through multiple methods:

* **Expert Validation**: Weights were reviewed by **social workers, NGOs, and child rights activists** to ensure alignment with real-world evidence.
* **Historical Case Analysis**: Trends were cross-verified with previously reported child marriage cases.
* **Internal Consistency Checks**: Scores were analyzed for variance and sensitivity to weight changes.

Additionally, a final **comparative analysis** was performed between Power BI dashboard insights and raw data trends to ensure consistency and reliability in the classification methodology.

## **3.2 Clustering Approach**

### **Feature Selection and Transformation**

Feature selection was guided by empirical research, prioritizing factors significantly impacting child marriage vulnerability. Variables such as Age (A)(A), Economic Instability (E)(E), Birth Registration (B)(B), Education (Ed)(Ed), and Social Disadvantage (S)(S) were selected due to their distinct contributions to risk assessment.

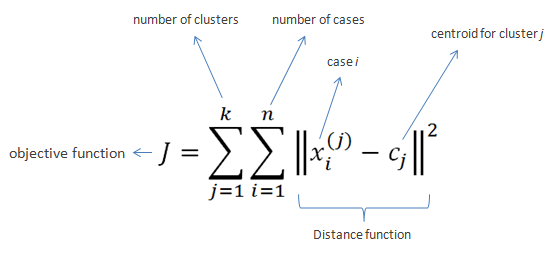
**Principal Component Analysis (PCA)** was then applied to reduce dimensionality while preserving variance, transforming the data as:

Z=WX

where Z represents the transformed dataset, W is the eigenvector matrix, and X is the original dataset.

### **Clustering Approach**

A clustering technique was employed to classify the dataset into distinct groups based on vulnerability levels. The chosen clustering algorithm minimizes the sum of squared distances within clusters, defined as:

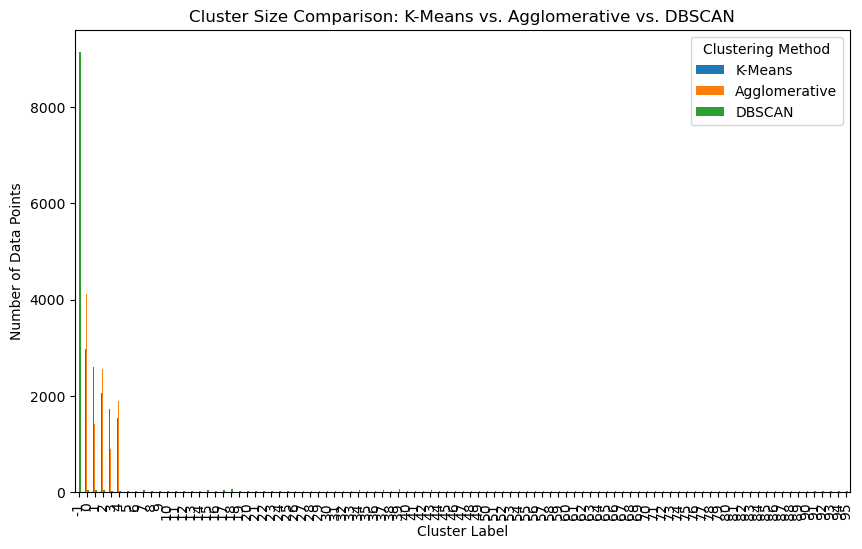
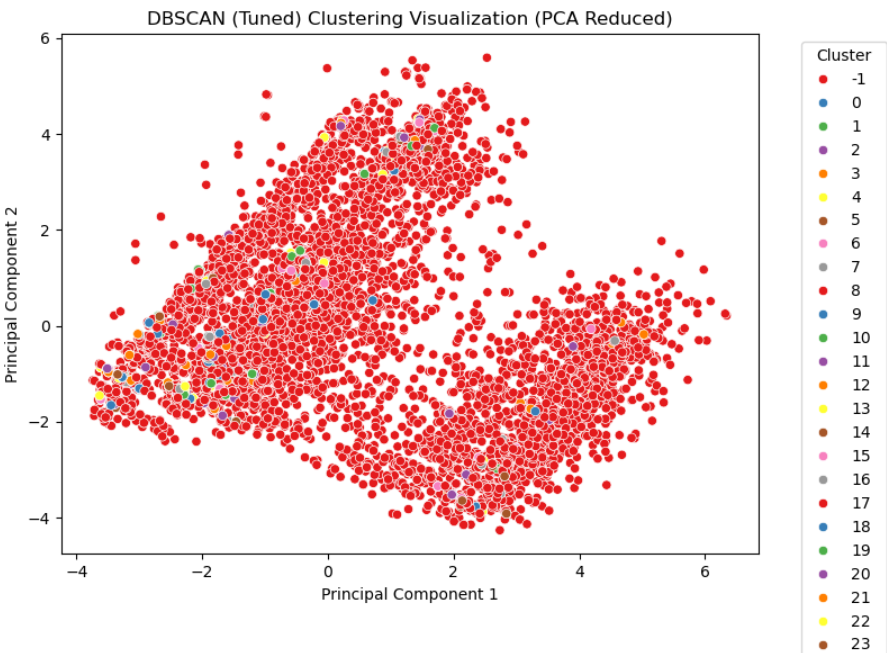


where xi represents data points, μj is the cluster centroid, and k is the number of clusters. The optimal number of clusters was determined using the elbow method by analyzing the within-cluster sum of squares (WCSS). The **K-Means++** initialization technique was used to optimize centroid placement, reducing convergence time and improving cluster accuracy.

**Why K-means?**

Based on a comparative analysis of clustering methods—K-Means, Agglomerative Clustering, and DBSCAN—K-Means emerges as the most effective technique for segmenting child marriage vulnerability groups. The PCA visualizations reveal that K-Means produces well-defined, compact clusters with clear boundaries, minimizing intra-cluster variance. In contrast, Agglomerative Clustering tends to merge smaller groups into a dominant cluster, reducing its capacity for nuanced segmentation, while DBSCAN proves overly aggressive, classifying the majority of points as noise and generating numerous small, incoherent clusters. The resulting structure from DBSCAN lacks interpretability, with unreliable silhouette scores due to extreme noise labeling, making it unsuitable for meaningful vulnerability analysis.

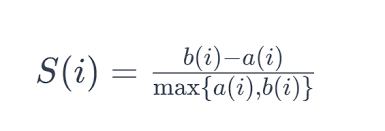
Furthermore, feature-based comparisons highlight K-Means' superior capacity to capture high-risk groups. It effectively isolates clusters with distinct proportions of unregistered children, differentiates economic vulnerability through occupation types, and distributes cluster sizes relatively evenly. Agglomerative Clustering, though useful for identifying hierarchical relationships, suffers from imbalanced cluster sizes and weaker segmentation of critical features like birth registration and ration card ownership. The dominance of a single large cluster further dilutes its effectiveness. Given these outcomes, K-Means offers the best balance of clarity, interpretability, and segmentation strength. Therefore, it is the most appropriate method for identifying distinct vulnerability groups, aligning both statistically and practically with the goal of targeted intervention.



### 

### **Cluster Analysis**

Post-clustering, each cluster was analyzed and classified into vulnerability levels based on statistical measures such as silhouette scores and inter-cluster distance calculations:



where a(i) is the average intra-cluster distance, and b(i) is the minimum average inter-cluster distance. The classification resulted in the following vulnerability levels:

* **Extreme Vulnerability**
* **High Vulnerability**
* **Moderate Vulnerability**

### **Machine Learning Models**

To further analyze and predict child marriage vulnerability, supervised machine learning models were applied.

#### 

#### **Random Forest Classifier**

The Random Forest model was employed for classification. This ensemble learning method operates by constructing multiple decision trees and aggregating their predictions. The final classification is based on majority voting:

ŷ = (1/M) ∑ T\_m(x)

where M is the total number of trees, and Tm(x)T\_m(x) represents the prediction of each tree.

#### **XGBoost Classifier:**

The XGBoost algorithm, an advanced gradient boosting technique, was also applied for prediction. It minimizes a regularized loss function:  
  
 L(θ) = ∑ l(ŷ\_i, y\_i) + ∑ Ω(f\_k)  
  
where l(ŷ\_i, y\_i) is the loss function (e.g., log loss for classification), and Ω(f\_k) is the regularization term to prevent overfitting.  
  
 **The optimization process** uses the following weight update in gradient boosting:  
  
 w^(t+1) = w^(t) - η ∂L/∂w  
  
 where η is the learning rate, and ∂L/∂w represents the gradient of the loss function.

### **Model Evaluation and Validation**

The models were evaluated using accuracy, precision, recall, and F1-score metrics:  
  
 Accuracy = (TP + TN) / (TP + TN + FP + FN)  
 Precision = TP / (TP + FP)  
 Recall = TP / (TP + FN)  
 F1-score = 2 × (Precision × Recall) / (Precision + Recall)  
  
where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives.  
  
To ensure model robustness, k-fold cross-validation was performed, where the dataset was split into k subsets, training on k-1 subsets and testing on the remaining one. The average performance was then computed to avoid overfitting.  
  
Hyperparameter tuning was conducted using Grid Search CV, optimizing parameters such as the number of trees in Random Forest and learning rate in XGBoost to enhance predictive accuracy. The final selected parameters were those yielding the highest cross-validation scores.  
  
This structured approach ensures a data-driven assessment of child marriage vulnerability, enhancing accuracy and interpretability through statistical rigor, clustering validation techniques, predictive modeling, and rigorous model evaluation.

# Limitations of the Study

Despite the robustness of the methodology employed in this study, several limitations must be acknowledged:

1. **Data Quality and Availability**

The accuracy and generalizability of the findings depend on the quality and completeness of the dataset used. Missing values were handled through imputation techniques, but potential biases arising from incomplete data cannot be entirely ruled out. Moreover, the study is constrained by the availability of region-specific data, which may limit the applicability of the model to broader or more diverse populations.

1. **Feature Selection Constraints**

The selection of features was based on empirical research and statistical significance. However, there may be latent variables influencing child marriage vulnerability that were not captured due to data limitations or the unavailability of certain socio-economic indicators. As a result, the model might not fully capture all the underlying causes of child marriage risk.

1. **Assumptions in Clustering and Classification**

The clustering approach relies on the assumption that vulnerability levels can be meaningfully segmented using statistical techniques. While methods such as the Elbow Method and Silhouette Score were used to validate cluster selection, human judgment and domain expertise still play a crucial role in interpreting these clusters. Additionally, classification models such as Random Forest and XGBoost assume patterns in the data remain consistent over time, which may not hold true in dynamic socio-economic environments.

1. **Model Generalization and Overfitting Risks**

Although cross-validation and hyperparameter tuning were conducted to improve model performance, there remains a risk of overfitting, particularly due to the complexity of the Random Forest and XGBoost models. The model's performance may degrade when applied to new datasets or regions with different socio-economic conditions.

1. **Ethical and Policy Considerations**

While the study aims to provide an objective, data-driven approach to assessing child marriage vulnerability, it does not replace the need for qualitative insights from affected communities. Furthermore, policy interventions based on the study should consider socio-cultural sensitivities and ethical concerns to avoid unintended consequences in implementation.

1. **Computational Constraints**

The methodology employed advanced machine learning techniques, including Principal Component Analysis (PCA), clustering, and ensemble learning models. While these methods enhance predictive accuracy, they also require significant computational resources. In resource-constrained settings, deploying such models at scale may pose challenges.

1. **Temporal and Spatial Limitations**

The study is based on a specific time frame, and the socio-economic factors influencing child marriage may evolve over time. Future studies should incorporate longitudinal data to track trends and assess the adaptability of the proposed methodology. Additionally, the study is geographically limited, and its findings may not be directly transferable to regions with significantly different socio-economic structures.

# Research Findings and Insights:

## Findings Based on Weighted Score Approach

#### **3.1 Vulnerability Scoring Analysis**

The weighted score approach provides a quantitative assessment of child marriage vulnerability across multiple socio-economic indicators. The analysis reveals that the **average vulnerability score** across all individuals is **0.63** (on a scale of 0–1), indicating a moderate to high level of risk. The standard deviation of **0.22** suggests some degree of variability in individual vulnerability levels.

A critical determinant of vulnerability is the prevalence of **households headed by minors** (children under 18 years), which accounts for **14.2%** of the total sample. This demographic exhibits significantly higher vulnerability scores, underscoring the heightened risk of early marriage among children who assume household responsibilities.

The **educational dimension** of vulnerability is reflected in the school regularity metric, where only **2.1% of children** report irregular attendance. Although this percentage appears relatively low, irregular school attendance and dropout rates are strong early indicators of child marriage risk, as education discontinuity often precedes forced or early marriages.

#### **3.2 Socio-Economic Factors and Vulnerability**

Economic conditions are a significant driver of child marriage vulnerability, as evidenced by **68.2% of families** falling under economically vulnerable categories based on **ration card scores** (average score = 0.68). Families experiencing financial distress are more likely to resort to early marriage as a coping mechanism, particularly for female children.

The **housing stability index** further reinforces economic vulnerability, with a score of **0.96** indicating that **96% of households** live in insecure housing conditions. Inadequate living environments correlate with increased risks of early marriage, as families in precarious housing situations often lack access to essential social services and support systems.

#### **3.3 Regional Trends and Risk Hotspots (Dashboard Insights)**

The geographical distribution of vulnerability scores reveals regional disparities in child marriage risk. Certain **districts in Chattisgarh** exhibit consistently high vulnerability scores, particularly in **rural and semi-urban** regions where socio-economic deprivation is more pronounced. In contrast, **urban areas** tend to report lower vulnerability scores; however, cases persist due to economic distress and entrenched socio-cultural practices.

A gender-based analysis highlights that **girls exhibit higher vulnerability scores than boys**, aligning with traditional gender norms and early marriage expectations. These findings suggest the need for **targeted interventions and awareness campaigns** focused on female child protection, particularly in high-risk regions.

#### **3.4 Risk Indicators and Predictive Patterns**

The study identifies key socio-cultural determinants that influence child marriage vulnerability. **Caste and religious affiliation** emerge as significant predictors, with certain groups displaying higher-than-average risk scores, indicating deep-rooted socio-cultural influences on early marriage practices.

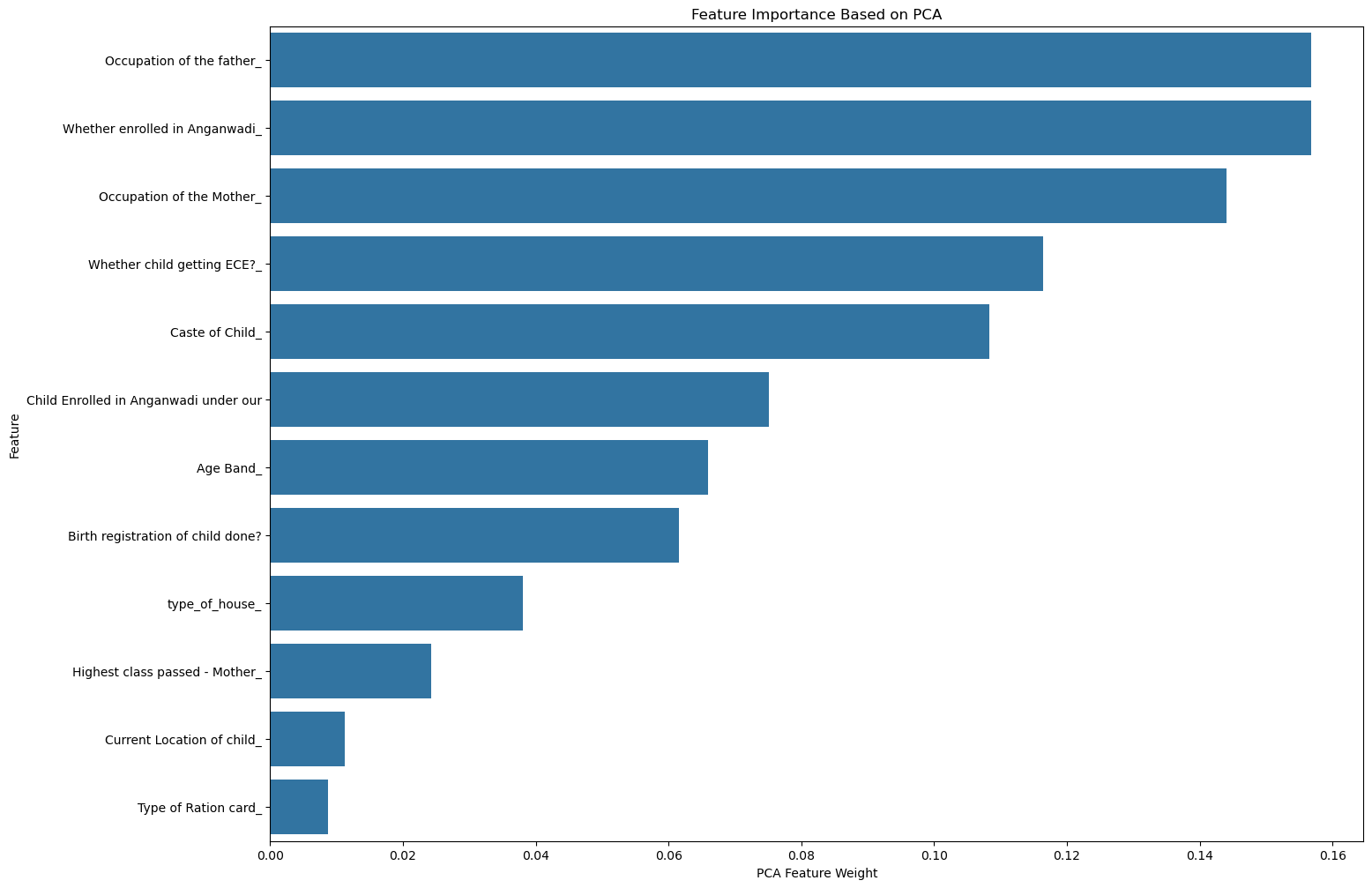
Family structure is another crucial determinant, where **father’s occupation and education level** directly correlate with child vulnerability scores. Children from **single-parent households or orphaned families** exhibit **substantially higher risk scores**, reinforcing the role of parental presence and financial stability in mitigating child marriage risks.

These findings provide a data-driven foundation for policy recommendations, emphasizing the necessity for **multi-dimensional interventions**, including economic support programs, education retention initiatives, and region-specific awareness campaigns to mitigate child marriage vulnerability effectively.

## **Analysis of Cluster Vulnerability for Child Marriage**

### **Analysis of Principal Component Analysis (PCA) Feature Importance**

The given bar chart represents the feature importance derived from **Principal Component Analysis (PCA)**. PCA is a dimensionality reduction technique that identifies the most significant features contributing to variance in the dataset. The horizontal axis represents the **PCA feature weight**, while the vertical axis lists the features ranked by their importance.



#### 

#### **Key Observations:**

1. **Most Influential Features:**
   * The **occupation of the father** has the highest importance in the principal components, suggesting a strong relationship between paternal employment and the underlying factors in the dataset.
   * **Enrollment in Anganwadi** follows closely, indicating that whether a child is part of an early childhood care program plays a crucial role in shaping the data’s variance.
   * **Occupation of the mother** is another significant factor, reflecting the role of maternal employment in influencing early childhood indicators.
2. **Moderately Significant Features:**
   * **Whether the child is receiving Early Childhood Education (ECE)** is a key determinant, highlighting its influence in the dataset.
   * **Caste of the child** emerges as a critical factor, likely reflecting socio-economic disparities in early childhood development.
   * **Enrollment in Anganwadi under a specific category** also carries considerable weight, further emphasizing the importance of government-supported early education initiatives.
3. **Lower Impact Features:**
   * **Age band and birth registration status** exhibit moderate importance, suggesting their relevance but comparatively lower contribution to overall variance.
   * **Type of house and mother's highest class passed** have relatively lower weights, indicating their lesser role in defining principal components.
   * **Current location of the child and type of ration card** have the least impact, suggesting that geographic and ration card-based socio-economic classifications contribute minimally to the dataset's variance.

#### **Implications of PCA Results:**

* The high importance of **parental occupation and Anganwadi enrollment** suggests that socio-economic factors and access to early education are dominant drivers in shaping the dataset.
* Policies aimed at improving **early childhood education** and **parental employment stability** may have a significant impact on outcomes associated with the dataset.
* The lower significance of **geographical location and ration card type** may imply that other socio-economic indicators hold greater predictive power.

## 

## **Analysis of Clusters:**

This study classifies clusters into five levels of vulnerability to child marriage: **Extreme Vulnerability**, **High Vulnerability**, **Moderate Vulnerability**, **Mild Vulnerability**, and **Low Vulnerability**. The classification is based on a systematic evaluation of key risk factors, including age distribution, parental occupation, maternal education, caste, birth registration, and Early Childhood Education (ECE) uptake. This analysis incorporates quantitative thresholds and comparative metrics to ensure precision and objectivity.

### **Analysis of Key Features by Clusters:**

**Rationale for Factor Selection**

The selection of the five factors—Age, Economic Instability, Birth Registration, Education, and Social Disadvantage—is grounded in established research identifying these variables as critical determinants of child marriage vulnerability. Each factor represents a distinct dimension of risk, and importantly, they correspond directly to specific columns within the dataset, ensuring both theoretical and empirical alignment.

**Justification for Exclusion of Other Factors**

The decision to exclude certain variables stems from the need to minimize redundancy. For instance, indicators such as house type and possession of a ration card often overlap with occupational status as proxies for poverty. Including all such variables would introduce multicollinearity without adding substantive value. Therefore, the selected factors offer a concise yet comprehensive representation of the key risks associated with child marriage.

**Empirical Validation**

To substantiate this selection, a variance analysis was conducted on the dataset. This statistical approach highlights the extent of variability each factor contributes, allowing us to distinguish those with meaningful influence from those with marginal impact. The following section presents the Python code and corresponding results, demonstrating how the chosen factors exhibit significant variance compared to the excluded ones.

#### **Age Distribution (11–18 Years) Across Clusters**

**Insights:** The age distribution across clusters reveals distinct patterns in the proportion of adolescents aged 11–14 and 15–18 years. Cluster 2 has the highest share of 11–14-year-olds (34%), indicating a younger demographic, while the 15–18 age group accounts for 19%, suggesting a notable presence of older adolescents at risk of child marriage. Clusters 0 and 4 show similar distributions, with around 32% of 11–14-year-olds and 20% and 17% of 15–18-year-olds, respectively, signaling a moderate risk. Cluster 3 has a lower proportion of older adolescents (16%), which may reflect stronger protective factors like school retention. Notably, Cluster 1 has the smallest adolescent population, with only 2% aged 11–14 and 1% aged 15–18, indicating limited immediate risk but a younger population base.

**Implications:** These insights highlight the need for cluster-specific interventions. Clusters 2, 0, and 4 should be prioritized for targeted programs addressing both early prevention for younger adolescents and immediate risk mitigation for older ones. Initiatives could include school retention strategies, legal education, and community awareness campaigns. Cluster 3 may benefit from reinforcing existing protective measures to sustain lower risk levels. While Cluster 1 shows minimal current risk, long-term preventive efforts are crucial to ensure that younger children remain safeguarded from future vulnerabilities. Tailored approaches will enhance the effectiveness of interventions, aligning them with the demographic realities of each cluster.

### 

#### **Birth Registrations “No” across clusters:**

### **Insights:** The data on birth registration across clusters reveals significant variation in unregistered births. Cluster 0 shows the highest proportion of individuals without birth registration at 40%, highlighting a critical gap in civil documentation. Cluster 2 follows with 23%, indicating moderate but still concerning levels of unregistered births. Cluster 3 records 16%, suggesting relatively better registration rates, though still requiring attention. Clusters 4 and 1 show the lowest rates, with 6% and 4%, respectively, pointing to more effective registration systems or better access to documentation in these areas. The disparity across clusters suggests uneven access to birth registration services.

### **Implications:** These findings underscore the urgent need for targeted interventions, particularly in Cluster 0, where the lack of birth registration may increase vulnerability to child rights violations, including early marriage and limited access to education and healthcare. Cluster 2 also warrants focused efforts to strengthen registration processes and public awareness campaigns. For Cluster 3, bolstering existing systems could further reduce the gap, while Clusters 1 and 4 should continue reinforcing their effective practices. Tailored strategies—such as mobile registration units, community outreach, and legal education—are essential to address the systemic barriers and ensure universal birth registration across all clusters.

### 

### 

#### **ECE Non-Uptake Across Clusters:**

**Insights:** The data on Early Childhood Education (ECE) non-uptake across clusters shows varying levels of disengagement. Cluster 1 has the highest proportion of non-uptake at 6%, indicating a significant gap in ECE participation. Cluster 3 follows at 4%, suggesting a moderate level of non-engagement. Cluster 4 records 2%, while Clusters 0 and 2 show the lowest levels, both at 1%, implying relatively better ECE inclusion in these areas. The differences between clusters highlight an uneven distribution of access to or participation in early education programs.

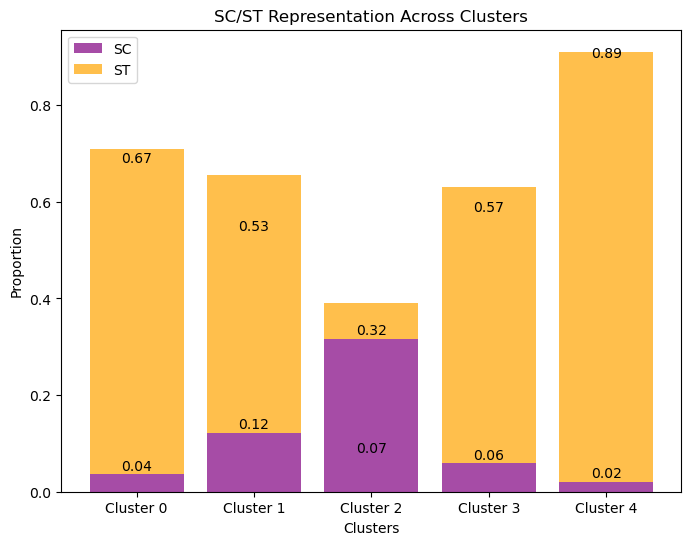
**Implications:** These disparities call for targeted policy responses to address barriers to ECE participation. Cluster 1 requires immediate intervention, potentially through community-based awareness programs, improved access to preschools, and support for families to overcome financial or logistical challenges. Cluster 3 would benefit from strengthened outreach efforts and infrastructure investments. Meanwhile, Clusters 0, 2, and 4 should focus on sustaining their relatively lower non-uptake rates by reinforcing current strategies. A cluster-specific approach is crucial to ensure equitable access to early education, which is foundational for long-term academic and social development.

### 

#### **SC/ST Representation Across Clusters:**

**Insights:** The SC/ST representation across clusters reveals notable variations. Cluster 4 has the highest proportion of ST representation at 89%, with minimal SC presence (2%), highlighting a strong concentration of ST individuals. Cluster 0 also shows a substantial ST proportion at 67%, while Cluster 1 follows at 53%. Cluster 3 has 57% ST representation and a small SC share of 6%. In contrast, Cluster 2 stands out with a relatively higher SC presence at 32%, the largest among all clusters, coupled with 7% ST representation. These differences underscore the uneven distribution of SC and ST populations across clusters.

**Implications:**The data suggests the need for cluster-specific strategies to ensure inclusive policymaking and targeted interventions. Clusters with a high ST presence, particularly Clusters 4 and 0, may benefit from culturally sensitive programs and improved access to resources tailored to tribal communities. Meanwhile, Cluster 2's relatively higher SC representation calls for a focus on their socio-economic empowerment and educational support. Policymakers must adopt a nuanced approach that considers the demographic composition of each cluster to promote equitable development and address community-specific challenges effectively.

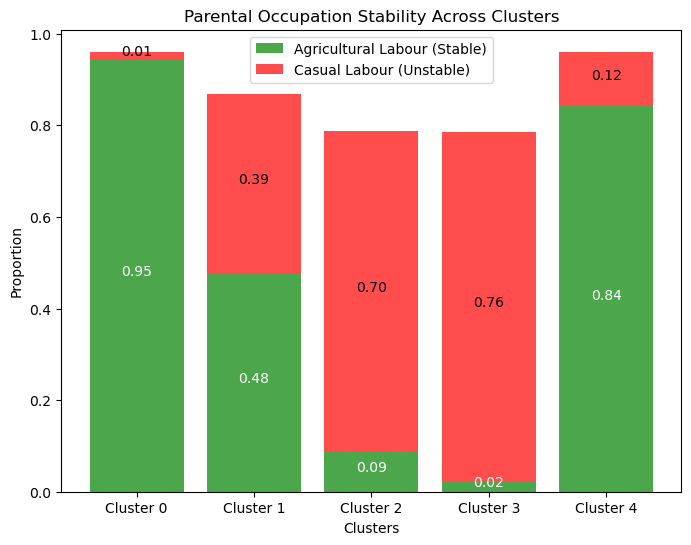
****

#### **Parental Occupation Stability Across Clusters:**

#### 

**Insights:**The chart highlights the distribution of parental occupation stability across clusters, distinguishing between agricultural labor (stable) and casual labor (unstable). Cluster 0 shows an overwhelming 95% of parents engaged in stable agricultural labor, with just 1% in casual labor. Cluster 4 also has a high proportion of stable jobs at 84%, though casual labor is slightly higher at 12%. In contrast, Clusters 2 and 3 are dominated by unstable casual labor, with 70% and 76% respectively, and minimal agricultural work (9% and 2%). Cluster 1 presents a more balanced split, with 48% in agricultural labor and 39% in casual labor.

**Implications:**These patterns suggest varying degrees of economic security across clusters. Clusters 0 and 4, with their higher levels of stable employment, may have relatively greater financial predictability, aiding in children's education and long-term planning. On the other hand, Clusters 2 and 3, where unstable casual labor dominates, likely face economic vulnerability, impacting children’s schooling and household stability. Targeted interventions should prioritize these clusters, focusing on skill development programs, income diversification opportunities, and social security schemes to enhance financial stability and reduce the risks associated with casual labor.



### **Key Risk Factors and Defined Thresholds**

To establish a rigorous framework, numerical thresholds were defined for critical indicators:

* **Lack of Birth Registration ("No")**:
  + High: >0.3
  + Moderate: 0.1–0.3
  + Low: <0.1
* **ECE Non-Uptake ("Not availing, services available")**:
  + High: >0.5
  + Moderate: 0.2–0.5
  + Low: <0.2
* **Maternal Illiteracy**:
  + High: >0.25
  + Moderate: 0.15–0.25
  + Low: <0.15
* **Proportion of Children Aged 11-18 Years**:
  + High: >0.5
  + Moderate: 0.3–0.5
  + Low: <0.3

These thresholds facilitate an objective assessment of each cluster's level of vulnerability.

### **Cluster Analysis with Quantitative Rankings**

The following is an in-depth cluster analysis, integrating occupation stability, quantitative comparisons, and ECE accessibility.

#### **Cluster 0**

* **Age Distribution**: Proportion aged 11-18 years: **0.523** (2nd highest, High)
* **Father’s Occupation**: Agricultural labor (0.945); though relatively stable, it remains susceptible to seasonal and market fluctuations.
* **Mother’s Occupation**: Agricultural labor (0.776) and unemployed (0.151), totaling **0.927** in vulnerable occupations.
* **Maternal Education**: Illiteracy: **0.240** (Moderate)
* **Caste**: Scheduled Caste (SC): **0.674** (High)
* **Birth Registration**: Lack of registration: **0.399** (Highest across clusters)
* **ECE Non-Uptake**: **0.856** (2nd highest)

**Summary**: High age-related risk, significant birth registration gaps, and substantial ECE non-uptake, exacerbated by partial occupation instability, position this cluster in the **High Vulnerability** category.

#### **Cluster 1**

* **Age Distribution**: Proportion aged 11-18 years: **0.032** (Lowest, Low)
* **Father’s Occupation**: Agricultural labor (0.479) and casual labor (0.390), totaling **0.869** in low-income jobs.
* **Mother’s Occupation**: Agricultural labor (0.359), casual labor (0.179), and unemployed (0.241), totaling **0.779**.
* **Maternal Education**: Illiteracy: **0.115** (Low)
* **Caste**: SC: **0.533** (Moderate)
* **Birth Registration**: Lack of registration: **0.039** (Lowest across clusters)
* **ECE Non-Uptake**: **0.059** (Lowest)

**Summary**: With a minimal proportion of at-risk ages, low rates of birth registration gaps, and strong ECE participation, this cluster is classified as having **Low Vulnerability**.

#### **Cluster 2**

* **Age Distribution**: Proportion aged 11-18 years: **0.529** (Highest, High)
* **Father’s Occupation**: Casual labor (0.699), indicating substantial economic instability.
* **Mother’s Occupation**: Unemployed (0.703), totaling **0.811** in vulnerable categories.
* **Maternal Education**: Illiteracy: **0.284** (High)
* **Caste**: SC: **0.495** (Moderate), Scheduled Tribe (ST): **0.317** (High)
* **Birth Registration**: Lack of registration: **0.228** (2nd highest, Moderate)
* **ECE Non-Uptake**: **0.946** (Highest)

**Summary**: Severe age risk, pronounced ECE non-uptake, and occupational instability categorize this cluster under **Extreme Vulnerability**.

#### **Cluster 3**

* **Age Distribution**: Proportion aged 11-18 years: **0.448** (Moderate)
* **Father’s Occupation**: Casual labor (0.765), contributing to economic precarity.
* **Mother’s Occupation**: Casual labor (0.632), offering some economic offset.
* **Maternal Education**: Illiteracy: **0.197** (Moderate)
* **Caste**: SC: **0.573** (Moderate)
* **Birth Registration**: Lack of registration: **0.159** (Moderate)
* **ECE Non-Uptake**: **0.010** (Low)

**Summary**: Moderate age risk coupled with maternal employment and strong ECE uptake classifies this cluster as **Mild Vulnerability**.

#### **Cluster 4**

* **Age Distribution**: Proportion aged 11-18 years: **0.489** (Moderate)
* **Father’s Occupation**: Agricultural labor (0.841), relatively stable but low-income.
* **Mother’s Occupation**: Agricultural labor (0.453), unemployed (0.136), totaling **0.589**.
* **Maternal Education**: Illiteracy: **0.348** (Highest)
* **Caste**: SC: **0.890** (Highest)
* **Birth Registration**: Lack of registration: **0.064** (Low)
* **ECE Non-Uptake**: **0.129** (Low)

**Summary**: Despite high maternal illiteracy and caste-based disadvantage, low birth registration gaps and moderate ECE uptake place this cluster in the **Moderate Vulnerability** category.

### **Summary of Key Metrics and Rankings**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metric** | **Cluster 0** | **Cluster 1** | **Cluster 2** | **Cluster 3** | **Cluster 4** |
| **11-18Y Proportion** | 0.523 (2) | 0.032 (5) | 0.529 (1) | 0.448 (4) | 0.489 (3) |
| **Lack of Birth Registration** | 0.399 (1) | 0.039 (5) | 0.228 (2) | 0.159 (3) | 0.064 (4) |
| **ECE Non-Uptake** | 0.856 (2) | 0.059 (5) | 0.946 (1) | 0.010 (4) | 0.129 (3) |
| **Maternal Illiteracy** | 0.240 (3) | 0.115 (5) | 0.284 (2) | 0.197 (4) | 0.348 (1) |
| **SC Caste Proportion** | 0.674 (2) | 0.533 (3) | 0.495 (4) | 0.573 (3) | 0.890 (1) |

### 

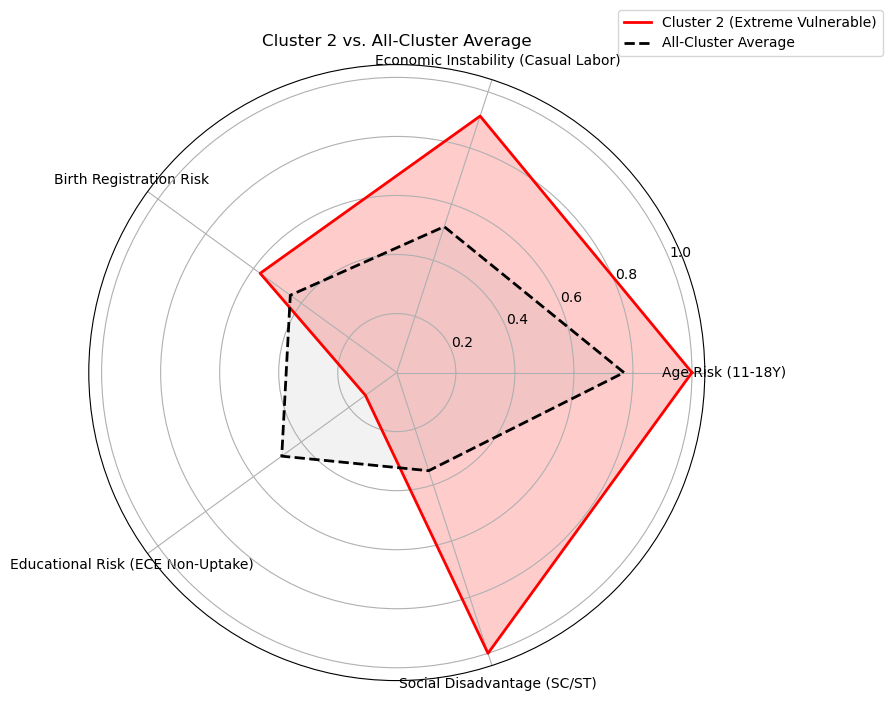
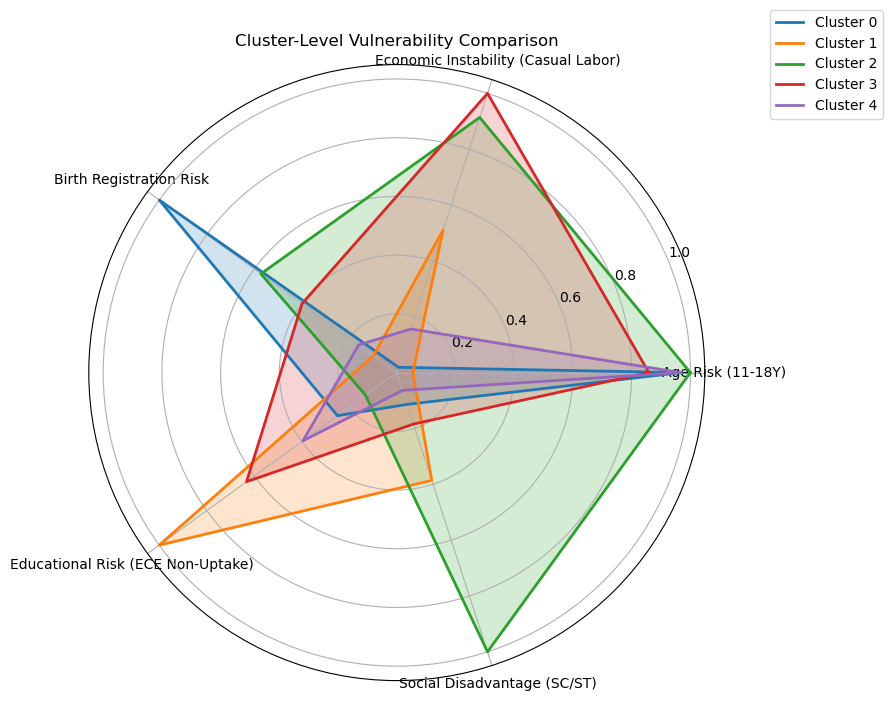
### **Cluster Vulnerability Levels**

1. **Cluster 2**: **Extreme Vulnerability**
2. **Cluster 0**: **High Vulnerability**
3. **Cluster 4**: **Moderate Vulnerability**
4. **Cluster 3**: **Mild Vulnerability**
5. **Cluster 1**: **Low Vulnerability**

**Alternative Cluster Names for Interpretability**

To enhance understanding, here are alternative names reflecting each cluster's defining characteristics:

1. **Cluster 2: "Older, Education-Deprived High-Risk Group"**
2. **Cluster 0: "Unregistered Agricultural Risk Group"**
3. **Cluster 4: "Service-Scarce Disadvantaged Group"**
4. **Cluster 3: "Working Parents Transitional Group"**
5. **Cluster 1: "Young Protected Low-Risk Group"**



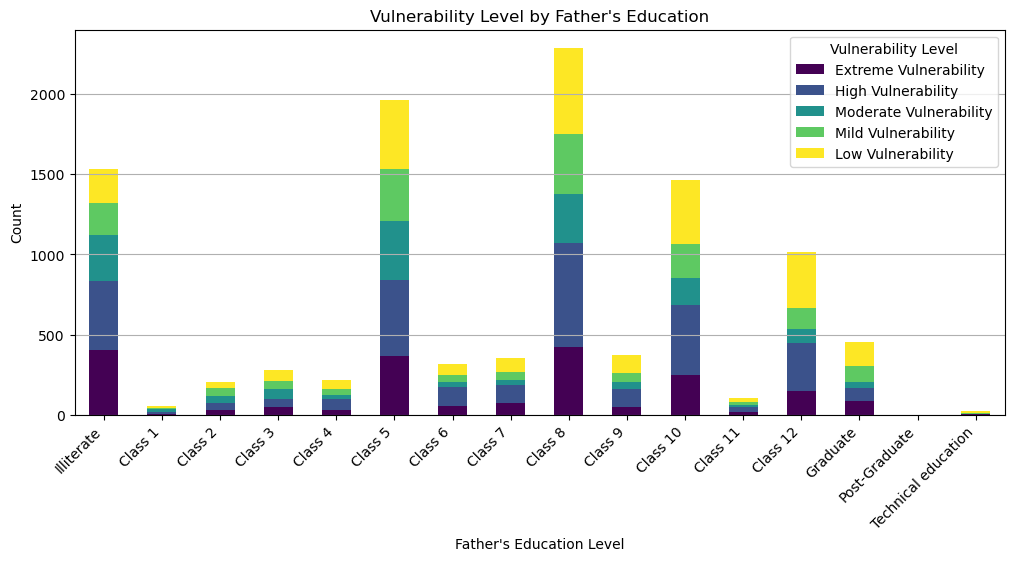
### **Analysis of Vulnerability by Key Factors and their implication:**

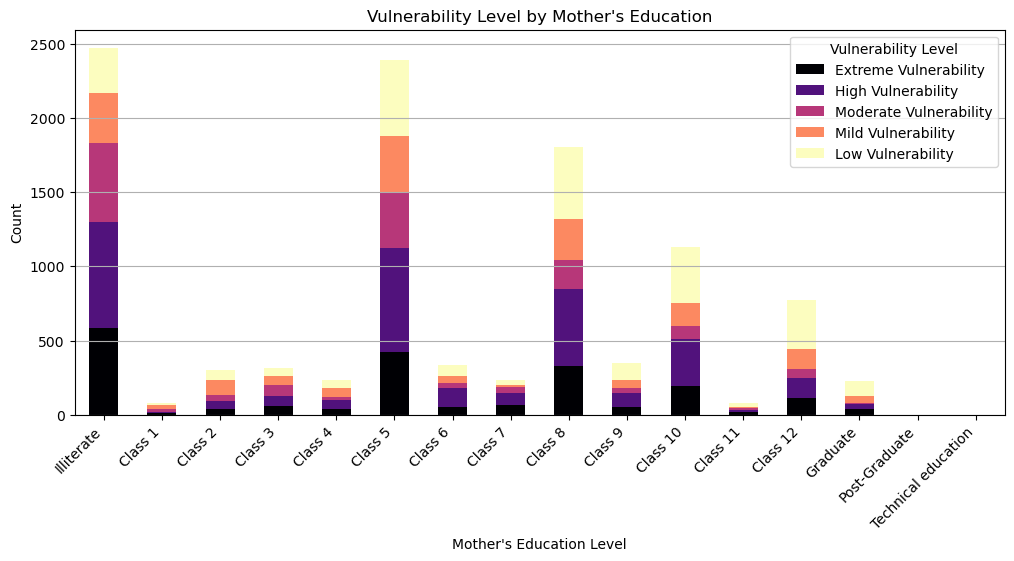
#### **Vulnerability level by Father’s and Mother’s Education**

1. **Father's Education and Vulnerability:**
   * Higher Vulnerability Among Illiterate Fathers: The highest counts of extreme and high vulnerability are concentrated among children whose fathers are illiterate, with a noticeable drop as education levels rise.
   * Class 5 and Class 8 Spikes: There's a sharp increase in vulnerability at Class 5 and Class 8 levels, which may suggest that fathers with incomplete primary or middle school education struggle with economic stability or access to resources.
   * Graduate and Post-Graduate Levels: Vulnerability levels drop significantly at higher education levels (graduate and above), with most falling into "low" or "mild" vulnerability categories.
2. **Mother's Education and Vulnerability:**
   * Illiteracy and Extreme Vulnerability: The trend is even more pronounced for mothers—children of illiterate mothers show the highest counts of extreme and high vulnerability.
   * Class 5 and Class 8 Patterns: Similar to fathers, Class 5 and Class 8 show spikes, implying that incomplete primary education for mothers contributes strongly to household instability.
   * Limited Higher Education Representation: There are very few mothers with post-graduate or technical education, indicating systemic barriers to female education and its long-term impact on family vulnerability.

### **Implications:**

1. Targeted Interventions for Illiterate Households:  
   * Given that extreme and high vulnerability levels are overwhelmingly linked to parental illiteracy, interventions should prioritize educational support and skill-building programs for these families.
   * Programs to improve adult literacy, especially for mothers, can have a cascading effect on household stability and children's well-being.
2. Focus on Early School Dropouts:  
   * The spikes at Class 5 and Class 8 suggest a dropout pattern at key transition points in education. Policies should address why parents may be leaving school early — whether due to economic hardship or societal pressures — and create targeted programs to retain children in school.
3. Gender-Specific Strategies:  
   * The stark contrast in the educational attainment of fathers versus mothers suggests a need for gender-sensitive interventions. Investing in female education, vocational training, and maternal empowerment can help break the cycle of vulnerability.
4. Higher Education Incentives:  
   * Since vulnerability drops dramatically for parents with higher education, there is a strong case for scholarships, financial aid, and community incentives to encourage school completion, especially for girls.

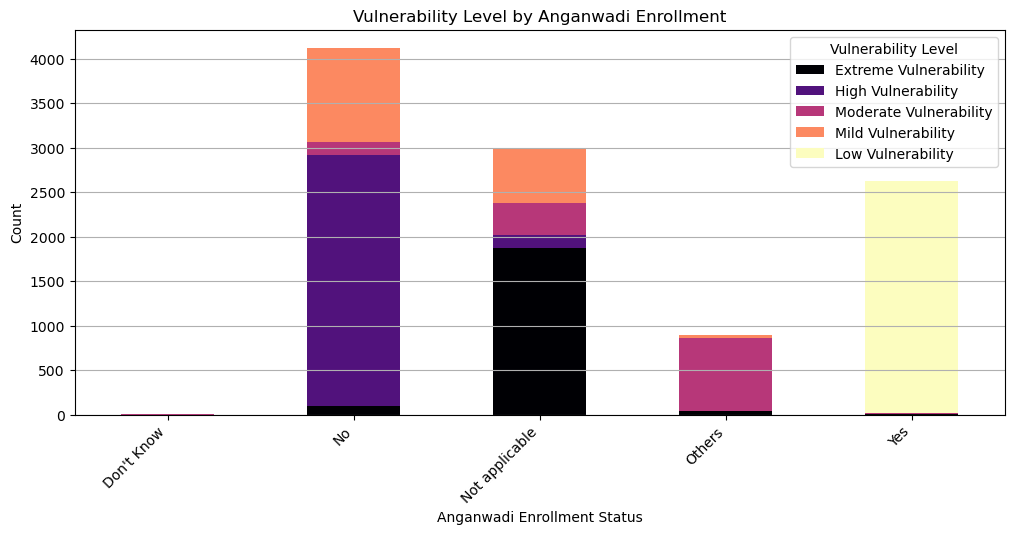




#### **Vulnerability Levels by Anganwadi Enrollment Status**

**Insights:** The graph reveals a clear correlation between Anganwadi enrollment and vulnerability levels. Children not enrolled in Anganwadi centers show significantly higher counts of extreme and high vulnerability, emphasizing the critical role these centers play in reducing risk. In contrast, children who are enrolled predominantly fall into the low vulnerability category, highlighting the protective impact of early childhood care programs. The "Not applicable" and "Others" categories also show moderate to mild vulnerability, suggesting gaps in service access or awareness. The minimal "Don't know" responses may reflect limited parental knowledge about Anganwadi services, pointing to an underlying issue of information dissemination.

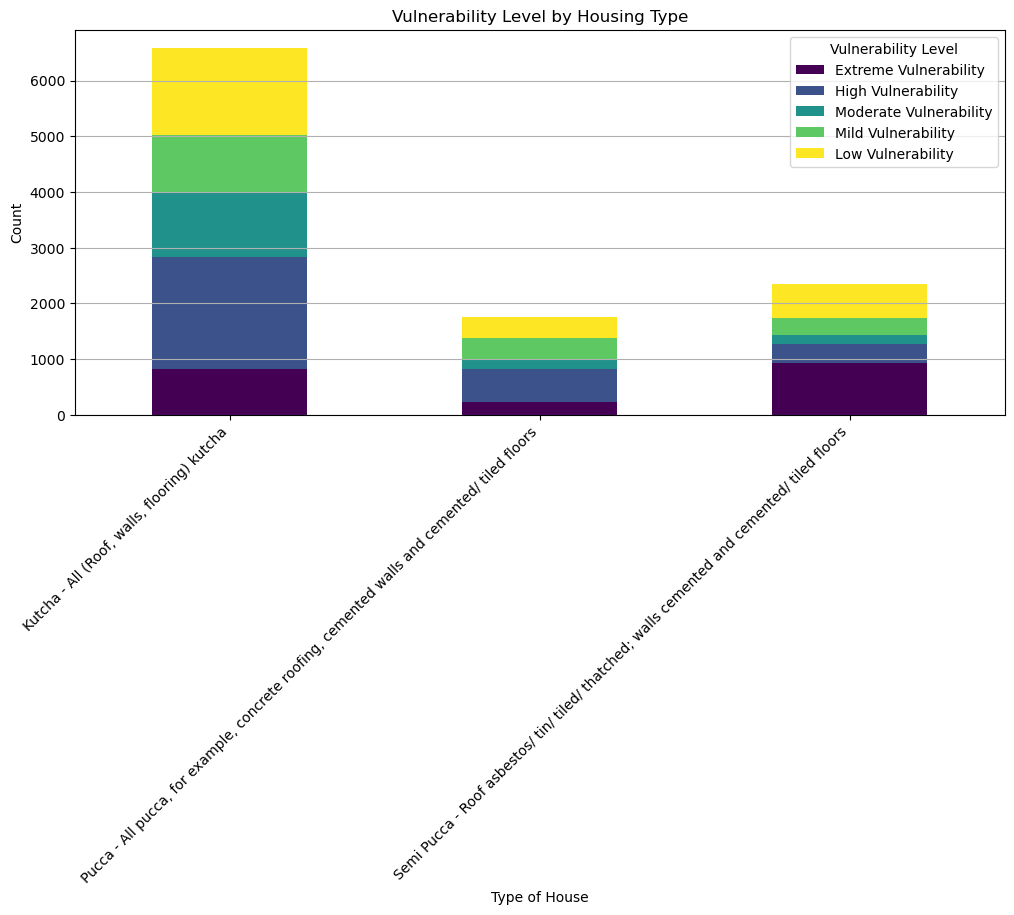
**Implications:** These findings underscore the need to strengthen Anganwadi outreach efforts, especially in communities with low enrollment, to bridge the gap between vulnerable households and essential services. Targeted campaigns should address barriers to enrollment — such as distance, social stigma, or lack of awareness — while also investigating why some children fall into ambiguous categories like "Not applicable." Policymakers must prioritize expanding Anganwadi coverage and enhancing parental education about early childhood programs. Additionally, robust data tracking systems can help identify high-risk areas, allowing for timely and evidence-based interventions to reduce child vulnerability effectively.

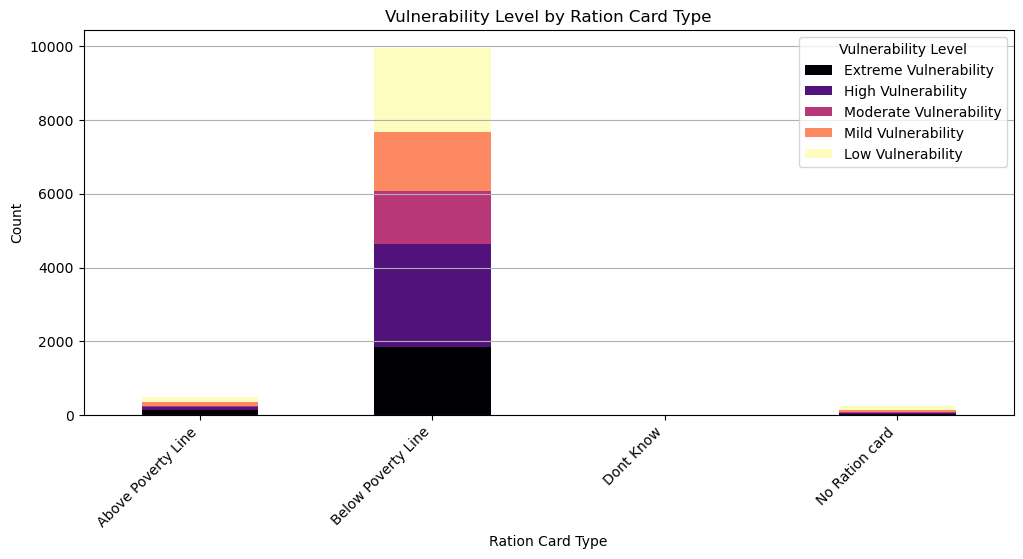


#### **Vulnerability Levels by Housing Types and Ration cards**

**Insights:** The visualizations reveal a strong correlation between socioeconomic indicators — housing type and ration card type — and vulnerability levels. Households residing in *kutcha* houses, characterized by non-durable materials for walls, roofs, and floors, exhibit significantly higher counts of extreme, high, and moderate vulnerability compared to those in *pucca* houses made of concrete and durable materials. Similarly, individuals holding Below Poverty Line (BPL) ration cards show markedly higher vulnerability levels, with extreme and high vulnerability categories dominating the distribution. In contrast, those with Above Poverty Line (APL) cards and those without ration cards report minimal vulnerability. Semi-pucca houses and BPL cardholders present a transitional pattern, reflecting moderate vulnerability. These findings highlight the intersection between inadequate housing conditions, economic precarity, and heightened vulnerability, emphasizing the compounded risk factors faced by marginalized populations.

**Implications:** These insights call for integrated policy interventions addressing both housing insecurity and economic hardship. Programs aimed at improving housing infrastructure should prioritize *kutcha* households by providing financial support, durable construction materials, and access to secure housing. Simultaneously, targeted welfare schemes must be reinforced for BPL cardholders, ensuring access to essential resources such as food, healthcare, and education. The overlap between housing type and ration card status suggests the need for a multi-sectoral approach, combining housing upgrades with strengthened social protection mechanisms. Furthermore, data-driven policy design should leverage these insights, ensuring resource allocation aligns with vulnerability levels and reaches those at greatest risk. Establishing robust monitoring systems will enhance accountability and responsiveness, fostering a more resilient and inclusive support framework.

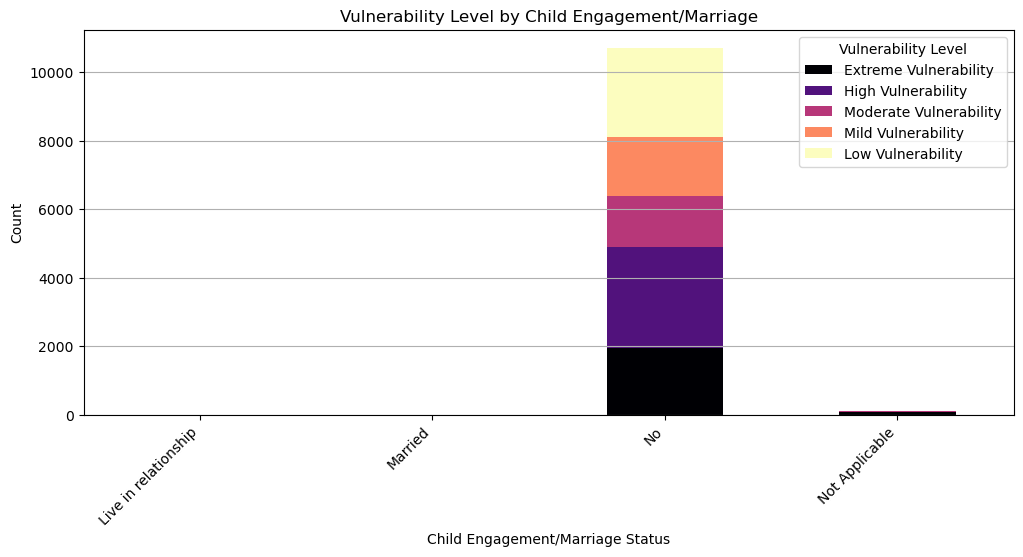




#### **Vulnerability Analysis by Child Engagement/Marriage:**

**Insights:** The visualization reveals a notable association between child engagement/marriage status and vulnerability levels. The majority of individuals categorized under "No" for child engagement/marriage display a wide distribution across all vulnerability levels, with a pronounced concentration in the extreme, high, and moderate categories. This suggests that children who are not engaged or married still experience considerable socioeconomic risks. Conversely, the "Married" and "Live-in relationship" categories show negligible counts, indicating that child marriage or engagement may be relatively rare in the dataset, though it remains crucial to acknowledge the intersectionality of vulnerabilities for those few cases. The "Not Applicable" group reflects minimal vulnerability, potentially representing individuals outside the age range for child engagement/marriage, further emphasizing the link between age, relationship status, and exposure to hardship.

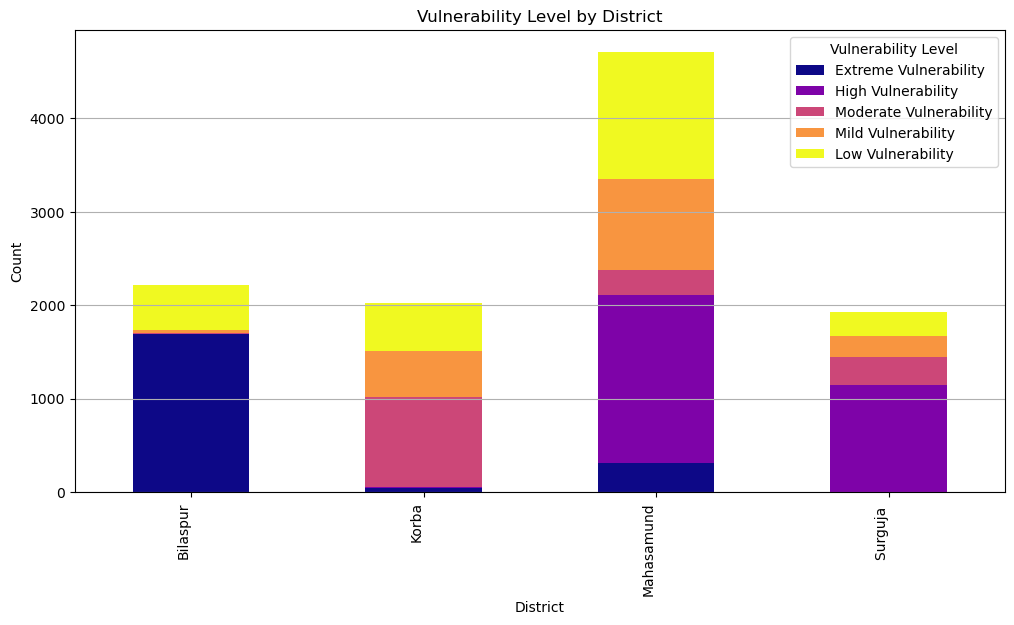
**Implications:** The findings underscore the need for targeted interventions aimed at reducing vulnerabilities among children who are not engaged or married, with a focus on addressing the root causes of extreme and high vulnerability, such as poverty, lack of education, and inadequate social protection. Policies should prioritize preventive strategies, such as strengthening education access, creating safe community spaces, and enhancing economic support for at-risk families to break the cycle of vulnerability. Although the data shows limited instances of child marriage or engagement, proactive measures should continue to enforce legal safeguards and community awareness campaigns to prevent these practices. Integrating these insights into broader child welfare programs will ensure a comprehensive approach, fostering both immediate relief and long-term resilience for vulnerable populations.

****

#### **Vulnerability Analysis by District:**

**Insights:** The visualization highlights the variation in vulnerability levels across districts, revealing that Mahasamund experiences the highest overall vulnerability, with significant counts in the extreme, high, and moderate categories. This suggests a concentration of socioeconomic challenges within this district. Bilaspur, while exhibiting a relatively high number of extreme vulnerability cases, shows less representation in the moderate to low categories, indicating a sharper divide. Korba reflects a more balanced distribution but with fewer cases of extreme vulnerability. Surguja, on the other hand, records the lowest overall counts, though high and moderate vulnerabilities remain prominent, pointing to persistent localized challenges. These patterns underscore the district-specific nature of vulnerability, suggesting differing underlying factors such as economic conditions, education levels, or access to resources.

**Implications:** The findings call for district-specific policy interventions, with Mahasamund requiring urgent, multifaceted strategies aimed at reducing extreme and high vulnerability—such as strengthening social welfare programs, enhancing livelihood opportunities, and expanding educational access. Bilaspur’s focus may benefit from targeted poverty alleviation programs, while Korba’s more even spread suggests a need for both preventive and remedial measures. For Surguja, consolidating existing support systems can help address moderate and high vulnerabilities before they escalate. Policymakers must adopt a decentralized approach, tailoring interventions to the unique socioeconomic conditions of each district to foster equitable development and resilience across regions.

****

### **Analysis of Random Forest and XGBoost Models in Predicting Child Marriage Vulnerability**

#### **1. Model Performance Evaluation**

**Random Forest Model:**

* The Random Forest model achieved an accuracy of 99.08%, indicating a highly effective predictive capacity.
* Class-wise analysis showed balanced performance, with precision, recall, and F1-scores mostly above 0.98.
* The lowest recall (96.9%) was observed for the “Moderate Vulnerability” category, suggesting some misclassification between moderate and neighboring categories.
* The perfect precision score (1.0) for “Low Vulnerability” indicates no misclassification of other classes into this category.
* The high macro and weighted averages imply robust generalization across classes. However, the near-perfect accuracy raises concerns about potential overfitting, especially if the dataset lacks sufficient diversity.

**XGBoost Model:**

* XGBoost also achieved 99.08% accuracy, with strong performance across all vulnerability levels.
* “High Vulnerability” had a perfect recall (100%), ensuring all high-risk children were correctly identified, making the model reliable for targeted interventions.
* The “Extreme Vulnerability” category had slightly lower precision (98.48%), suggesting minor misclassification into adjacent categories.
* “Moderate Vulnerability” had the lowest recall (96.61%), indicating overlaps with mild and high-risk groups, highlighting a need for more granular feature refinement.
* The model’s 99.90% F1-score for “Low Vulnerability” shows high confidence in predicting low-risk children.

#### **2. Feature Importance and Key Influencing Factors**

**Random Forest Insights:**

* Primary Drivers:
  + *Lack of Anganwadi Enrollment* (34.31%) emerged as the most significant predictor of child marriage vulnerability, emphasizing the protective role of early childhood education and social engagement.
  + *Lack of Early Childhood Education (ECE)* (18.74%) was the second strongest factor, linking early educational interventions to delayed marriages.
* Economic Influences:
  + *Father’s Occupation* (13.02%) and *Mother’s Occupation* (9.83%) showed that financial instability is linked to early marriage decisions, reflecting economic pressures.
* Protective Factors:
  + *Enrollment in Anganwadi under an Intervention Area* (8.64%) highlighted the positive impact of targeted interventions.
* Moderate Influences:
  + *Caste of Child* (4.62%), *Age Band* (2.87%), and *Birth Registration* (2.48%) moderately affected vulnerability.
* Least Influential Factors:
  + *Housing Type* (0.79%), *Current Location of Child* (0.52%), *Mother’s Education* (0.32%), and *Ration Card Type* (0.00%) were minimal predictors, indicating economic and educational factors outweighed these structural variables.

**XGBoost Insights:**

* Strongest Predictors:
  + *Anganwadi Enrollment* (54.10%) was the most influential factor, reinforcing its protective role.
  + *Early Childhood Education (ECE)* (23.47%) followed closely, underlining the importance of structured early education.
* Moderate Predictors:
  + *Mother’s Occupation* (6.16%) and *Father’s Occupation* (6.11%) showed economic stability's moderate impact.
* Lower Influences:
  + *Caste of Child* (1.70%) and *Birth Registration* (1.07%) had marginal roles.
* Least Impactful Factors:
  + *Age Band* (0.34%), *Housing Type* (0.27%), *Current Location of Child* (0.10%), *Mother’s Education* (0.09%), and *Ration Card Type* (0.00%) demonstrated minimal predictive power.

#### 

#### **3. Policy Implications and Key Takeaways**

* Education Access: Both models highlight *Anganwadi Enrollment* and *ECE* as the most crucial protective factors against child marriage. Expanding these programs can significantly mitigate risk.
* Economic Stability: Parents’ occupations reflect financial stress as a key driver. Supporting families with financial aid and employment programs may delay early marriage decisions.
* Legal Identity and Social Factors: While caste and birth registration moderately influence vulnerability, their effects are secondary to educational and economic factors. Ensuring universal birth registration can aid in enforcing anti-child-marriage laws.
* Minimally Relevant Factors: Housing conditions, ration card status, and a child’s current location have limited impact. Therefore, intervention strategies should focus more on economic stability and education rather than infrastructure alone.

# Conclusion & Recommendation:

This study provides a comprehensive analysis of child marriage vulnerability in Chattisgarh using statistical modeling and machine learning. The findings highlight the influence of socioeconomic and educational factors, with Anganwadi enrollment, ECE, and parental occupation stability emerging as key predictors. The clustering analysis revealed stark disparities across risk levels, reinforcing the need for targeted interventions. The high accuracy of the Random Forest and XGBoost models validates the methodology’s reliability for identifying and addressing child marriage risks.

While national child marriage rates have declined, regional disparities persist, driven by poverty, gender norms, and limited educational access. A one-size-fits-all approach is inadequate; tailored strategies rooted in data-driven insights are crucial.

**Recommendations:**

1. Expand Anganwadi and ECE programs to ensure equitable access in high-risk regions.
2. Introduce vocational training and microfinance schemes to enhance family economic stability.
3. Implement scholarships and school retention programs, particularly for girls.
4. Organize birth registration drives to ensure legal identity and protect child rights.
5. Establish a centralized data system to track child marriage vulnerability indicators.
6. Partner with NGOs for community-based advocacy to challenge patriarchal norms.
7. Foster cross-sector collaboration between education, health, and social welfare departments.