**PROJECT TITLE:**

Assessment of Marginal Workers in Tamil Nadu-ADS

**PROBLEM DEFINITION:**

* H0e term "marginal" as used in economics. It is often stated that "economics happens on the margins". And this is an important distinction that most people just don't get. And this lack of getting this has serious implications in the discussion of economic policy. It can be thought of in more colloquial terms as "fence sitting", regarding some deciding parameter, usually in reference to changes in price

**PRE-PROCESSING:**

**STEPS:**

1. **DATA CLEANING**
2. **HANDLE MISSING VALUES**
3. **CATEGORICAL TO NUMERICAL REPRESENTATIONS.**

**DATA CLEANING:**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

**HANDLE MISSING VALUES:**

1. Deleting Rows with missing values
2. Impute missing values for continuous variable
3. Impute missing values for categorical variable
4. Other Imputation Methods
5. Using Algorithms that support missing values
6. Prediction of missing values
7. Imputation using Deep Learning Library — Datawig

**CATEGORICAL TO NUMERICAL REPRESENTATIONS:**

1. cat.codes Attribute
2. replace
3. Label Encoder

**ALGORITHM:**

1. **Purpose and Objectives**
2. **Data Quality and Bias**
3. **Fairness and Equity**
4. **Transparency and Explainability**
5. **Human Oversight**

**Purpose and Objectives**:

Begin by understanding the purpose and objectives of the algorithm. Is it designed to identify marginalized workers for support, allocate resources, or make decisions related to employment? Knowing the intended outcomes is crucial for assessing its effectiveness.

**Data Quality and Bias**:

Check the data sources and quality. Biases in data can lead to unfair outcomes, especially for marginalized groups. It's important to assess if the algorithm's training data is representative and if steps were taken to mitigate bias.

**Fairness and Equity**:

Evaluate the algorithm's fairness. Use fairness metrics like disparate impact, equal opportunity, and demographic parity to assess whether it disproportionately impacts marginalized workers.

**Transparency and Explainability**:

Assess the transparency and explainability of the algorithm. Can the decision-making process be understood and explained to those affected by it? Transparency is crucial for accountability.

**Human Oversight**:

Determine whether there is a human oversight component. While algorithms can assist decision-making, human judgment is often necessary, especially in complex or sensitive situations.

**PROJECT WORKFLOW:**

