**IBM NAANMUDHALVAN**

**APPLIED DATASCIENCE-PHASE 2**

**ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU**

**INNOVATION:**

**Step 1- Data Source:**

* We will begin by identifying publicly available datasets related to marginal workers. We will check official government websites, labor department databases, or research institutions for relevant data. For example, we can explore sources like the National Sample Survey (NSSO) or Census data.

**Step 2**- **Data Acquisition:**

* We will download or obtain the dataset, making sure it includes information on age, gender, and their respective industries. We may need to merge multiple datasets to compile this information.

**Step 3**- **Data Cleaning:**

* We will perform data cleaning to address issues such as missing values, outliers, and inconsistencies. Here's a general outline of what we will do:

**Handle Missing Data:** We will identify missing values and decide whether to impute them or remove rows/columns with excessive missing data. We will clearly document the approach used.

**Outlier Detection:** We will detect and decide how to handle outliers, which could distort our analysis. Common methods include z-scores or visual inspection.

**Consistency Check:** We will examine the data for inconsistencies in terms of data types, formatting, or illogical values. We will correct any discrepancies.

**Step 4- Principal Component Analysis (PCA):**

* Apply Principal Component Analysis (PCA), a machine learning technique, to simplify and extract meaningful patterns from the data.
* Use PCA to identify key variables and assess their influence on data variance.
* Visualize the results of PCA through bi-plots to enhance our understanding of how age, gender, and industrial work interrelate within the marginal worker demographics.

**Step 5- Interactive Dashboards:**

* Create interactive dashboards using data visualization tools such as Tableau or Power BI.
* Incorporate the findings from PCA into these dashboards to provide dynamic insights into age distribution, gender ratios, and industrial categorization simultaneously.
* Ensure that these dashboards are user-friendly and intuitive, allowing policymakers and stakeholders to explore the data effectively.

**FLOWCHART:**

    Begin Analysis

  Data Collection

 Data Refinement

 Data Examination

  Data Interpretation

Visual Data Display

Conclusion & End