# **Tamil Nadu Marginal Workers**

# **Innovation**

**Phase :02 Project Number : 03**

**Project Definition:**

This assessment focuses on analyzing the data from Tamil Nadu Marginal workers are those who, due to the nature of their employment, have limited access to stable employment and often face socioeconomic challenges. Understanding their demographic distribution and characteristics is crucial for policy-making and addressing socio-economic disparities.

**Project Objectives:**

* Data Extraction for the TN-Marginal Workers
* Cognos analytics in action for Tamil Nadu Marginal workers
* Conducting clustering analysis to identify patterns among different industrial categories and age groups.

**Data Set:**

Here we are using the data set from the govt organization official site(tn.data.gov.in).

This dataset from: <https://tn.data.gov.in/resource/marginal-workers-classified-age-industrial-category-and-sex-scheduled-caste-2011-tamil>

This dataset contains 71 columns and 595 rows.

**Data Extraction:**

Data extraction is a fundamental step in the process of data analysis using IBM Cognos

* Data Source Selection:

Gain a deep understanding of the problem statement and the context of marginal workers in Tamil Nadu. Cognos can connect to a wide range of data sources, including databases, spreadsheets, and web services. You'll need to configure data connections to these sources within Cognos.

* Data Modeling:

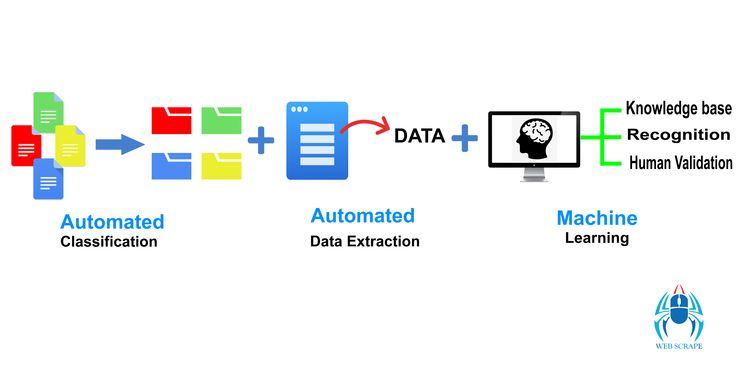
In Cognos, you can create a data model to define how the data is structured and organized. This involves identifying the tables, views, and relationships between data sources. This step helps ensure that the extracted data is in a format suitable for analysis.

* Query Building:

Once your data model is set up, you can build queries to extract the specific data you need. Cognos provides a visual query builder interface to help you design and customize your data queries.

* Data Extraction:

After defining your query, you can execute it to extract data from the selected sources. The extracted data can be in the form of tables or datasets, depending on the type of analysis you want to perform.



* Data Transformation:

In many data analysis scenarios, you might need to clean, transform, or reshape the extracted data. Cognos offers tools for data transformation and cleansing to ensure that the data is accurate and suitable for analysis.

* Data Integration:

If you're working with data from multiple sources, you can integrate and combine data from various sources into a unified dataset. This is especially important for comprehensive analysis.

* Data Storage:

Cognos typically doesn't store the extracted data permanently, but you can save the results of your analysis, reports, or dashboards for future reference.

* Analysis and Reporting:

Cognos typically doesn't store the extracted data permanently, but you can save the results of your analysis, reports, or dashboards for future reference

* Data Refresh and Automation:

Depending on your needs, you may want to set up automated data extraction and analysis processes to keep your data up to date. Cognos can schedule data refreshes and automate report generation

* Deployment and Sharing:

Finally, you can deploy your analysis results and share them with relevant stakeholders, which can be done through Cognos reports, dashboards, or other means.

**Analytics in Action:**

This data is often structured and organized to support data analysis and decision-making. Analytical data in Cognos can come from various sources and be transformed and modeled to meet the specific needs of business users. Here are key aspects related to analytical data in Cognos

* Data Sources:

Analytical data can originate from various sources, including databases, data warehouses, spreadsheets, and external data feeds. Cognos can connect to these sources to access and analyze data.

* Data Modeling:

Data modeling in Cognos involves defining the structure and relationships within the data. It helps users understand how different data elements are related and organized, making it easier to query and analyze the data.

* ETL (Extract, Transform, Load):

Data extraction, transformation, and loading are crucial processes to prepare data for analysis. Cognos provides tools and capabilities to extract data from source systems, transform it to suit analytical needs, and load it into its data model.

* Data Exploration:

Cognos offers features for data exploration, allowing users to navigate, filter, and interact with the data to identify patterns, trends, and anomalies.

* Data Visualization:

Cognos enables users to create various data visualizations, such as charts, graphs, and dashboards, to represent data in a visually meaningful way.

* Reporting:

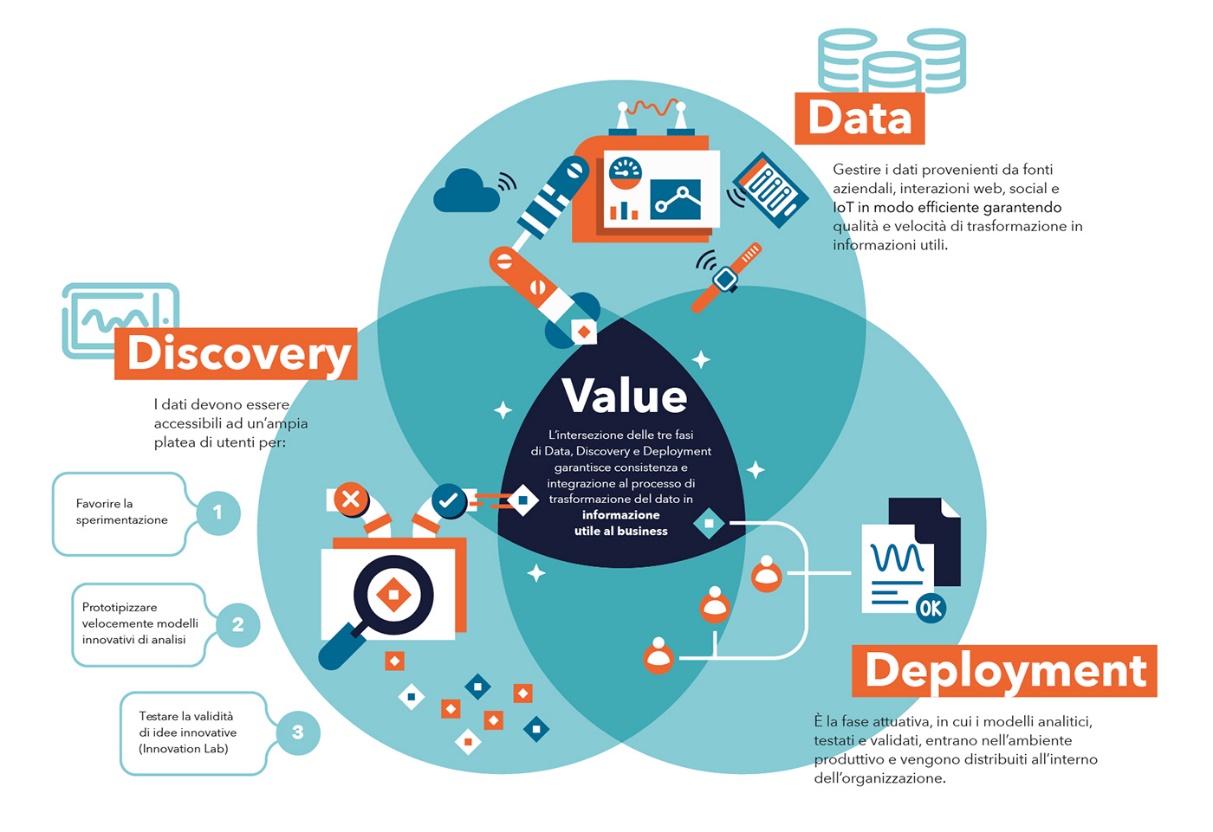
Cognos allows users to create reports based on the analytical data. These reports can be customized, scheduled, and distributed to relevant stakeholders.

* Ad-Hoc Analysis:

Business users can perform ad-hoc analysis by creating their own queries and reports, enabling them to explore data and generate insights without relying on IT or data analysts.

* Data Security and Governance:

Cognos provides tools for managing data security and governance. This ensures that only authorized users have access to sensitive data, and data remains compliant with organizational policies and regulations.



* Data Integration:

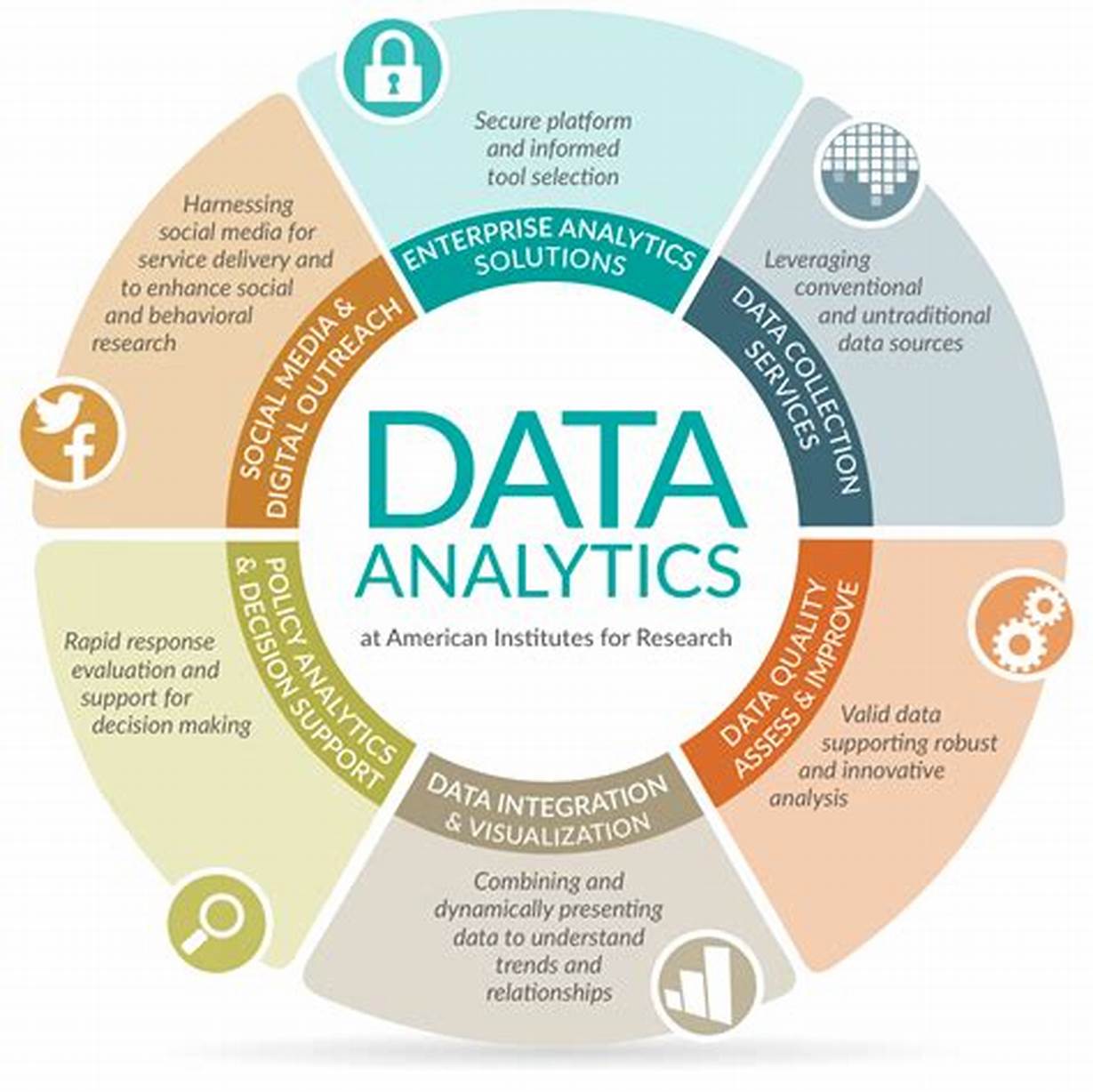
Analytical data in Cognos may involve the integration of data from various sources, making it possible to analyze and report on a comprehensive set of data.

* Data Performance:

Cognos optimizes data retrieval and analysis performance, ensuring that analytical processes run efficiently, even with large datasets.

**Analytics of Data:**

* Data Source Connection: Begin by connecting to your data sources. Cognos can connect to various data sources, including databases, spreadsheets, and web services. Configure the connections to access your data.
* Data Modeling: Create a data model within Cognos to define how your data is structured and organized. This typically involves identifying tables, views, and relationships between data sources. A well-defined data model makes it easier to work with the data.
* Query Building: Use Cognos Query Studio or Report Studio to create data queries. These queries allow you to extract specific data from your sources and define the information you need for analysis.
* Data Extraction: Execute your queries to extract the data from the sources. The extracted data can be in the form of tables or datasets.
* Data Transformation: Data may need to be cleaned, transformed, or reshaped for analysis. Cognos provides tools for data transformation and cleansing to ensure data accuracy.
* Data Integration: If your analysis requires data from multiple sources, you can integrate and combine data from various sources into a unified dataset.



* Analysis Tools: Cognos offers various tools for data analysis, including reporting, data exploration, and ad-hoc analysis. These tools allow you to identify patterns, trends, and insights in your data.
* Data Visualization: Create data visualizations, such as charts, graphs, and dashboards, to present your findings in a visually meaningful way.
* Reporting: Generate reports based on your analysis. Cognos reports can be customized, scheduled, and distributed to relevant stakeholders.
* Self-Service Analytics: Cognos offers self-service analytics capabilities, enabling business users to explore and analyze data independently, reducing the reliance on IT for every analysis request.
* Data Governance and Security: Manage data security and governance in Cognos to ensure that data remains compliant with organizational policies and regulations. Define who has access to what data.
* Performance Optimization: Cognos optimizes data retrieval and analysis performance, ensuring that analytical processes run efficiently, even with large datasets.
* Data Exploration and Drill-Down: Explore data interactively, apply filters, and drill down into details to uncover deeper insights.
* Data Export and Sharing: Export your analysis results, share insights, and collaborate with team members using Cognos's sharing and collaboration features.
* Automated Data Updates: Set up automated data refreshes to keep your analysis up to date.
* Deployment: Deploy your analysis results, reports, and dashboards for use within your organization.

**Visualization Selection:**

## **Bar Graphs**

Bar graphs are used show the distribution of qualitative (categorical) data.It shows the **frequency** of values in the data. Frequency is the amount of times that value appeared in the data.

Each category is represented with a bar. The height of the bar represents the frequency of values from that category in the data.

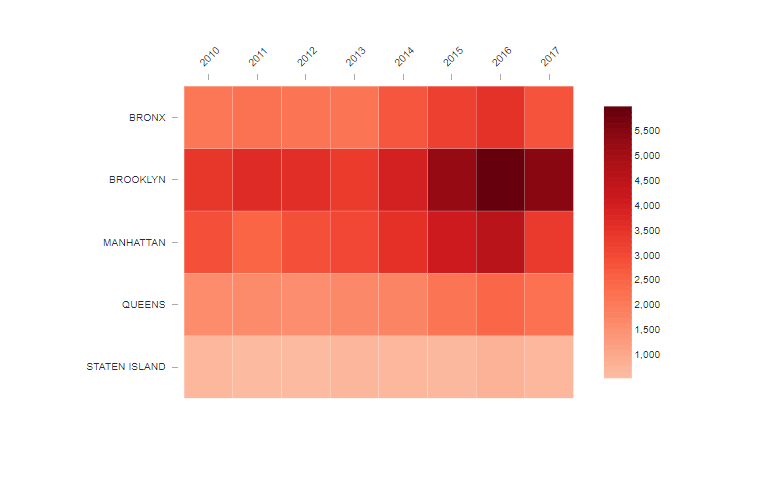
## **Pie Charts**

Pie graphs are used to show the distribution of qualitative (categorical) data .It shows the **frequency** or **relative frequency** of values in the data .Frequency is the amount of times that value appeared in the data. Relative frequency is the percentage of the total. Each category is represented with a slice in the 'pie' (circle). The size of each slice represents the frequency of values from that category in the data.

## **Heat Map**

A heat map is a two-dimensional representation of data in which various values are represented by colors. A simple heat map provides an immediate visual summary of information across two axes, allowing users to quickly grasp the most important or relevant datapoints. More elaborate heat maps allow the viewer to understand complex datasets.

A heat map is a way to represent data points in a data set in a visual manner. All heat maps share one thing in common -- they use different colors or different shades of the same color to represent different values and to communicate the relationships that may exist between the variables plotted on the x-axis and y-axis. Usually, a darker color or shade represents a higher or greater quantity of the value being represented in the heat map.



**Analyzing and Handling Data set:**

* Load the Dataset:
* Import the necessary libraries (e.g., Pandas, NumPy) in Python.
* Load your dataset into a data structure (e.g., a Pandas DataFrame)

import pandas as pd

data = pd.read\_csv('your\_dataset.csv')

* Missing Values:

Identify and handle missing values in your dataset. You can drop rows with missing values, impute values, or use other strategies depending on the context.

print(data.head())

print(data.info())

print(data.describe())

* Duplicate Data:

Check for and remove duplicate rows if necessary.

data.dropna() # To remove rows with missing values

data.fillna(value) # To fill missing values with a specific value

* Outliers:

Identify and handle outliers in your data. You can use statistical methods or visualization techniques to detect outliers.

# Example using Z-score for outlier detection

from scipy import stats

z\_scores = stats.zscore(data)

data\_no\_outliers = data[(z\_scores < 3).all(axis=1)]

* Data Transformation:

Transform data if needed, such as converting data types, encoding categorical variables, or scaling numerical features.

# Example: Encoding categorical variables

data = pd.get\_dummies(data, columns=['categorical\_column'])

* Feature Engineering:

Create new features from existing ones to improve model performance.

data['new\_feature'] = data['feature1'] + data['feature2']

* Data Splitting:

Split your dataset into training, validation, and test sets for model development and evaluation.

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

* Scaling and Normalization:

Scale or normalize numerical features to ensure that they have a similar scale and distribution.

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

* Data Analysis:

Perform the specific analysis you need, such as building machine learning models, statistical analysis, or hypothesis testing.

**Conclusion:**

In this phase, we have defined the problem, objectives, and outlined the design thinking process for the project. The ultimate goal is to provide meaningful insights that can inform policies and initiatives aimed at improving the lives of marginal workers in Tamil Nadu. Document your entire analysis process, including data sources, methods, and assumptions. This documentation is crucial for transparency and reproducibility.