**ASSESSMENT OF MARGINAL WORKERS IN TAMILNADU - A SOCIOECONOMIC ANALYSIS**

**INTRODUCTION:**

The analysis of the topic comprises examining the age, sexual orientation, and population demographics based on industry classification of workers who are marginalized in Tamil Nadu. The main goal is to complete an economic and social analysis and provide examples of how marginal employees are distributed across various demographic groups. The goals of this project have been outlined, a data analysis plan has been developed, useful visualization types have been identified, and Python and data visualization libraries are being used for the analysis.

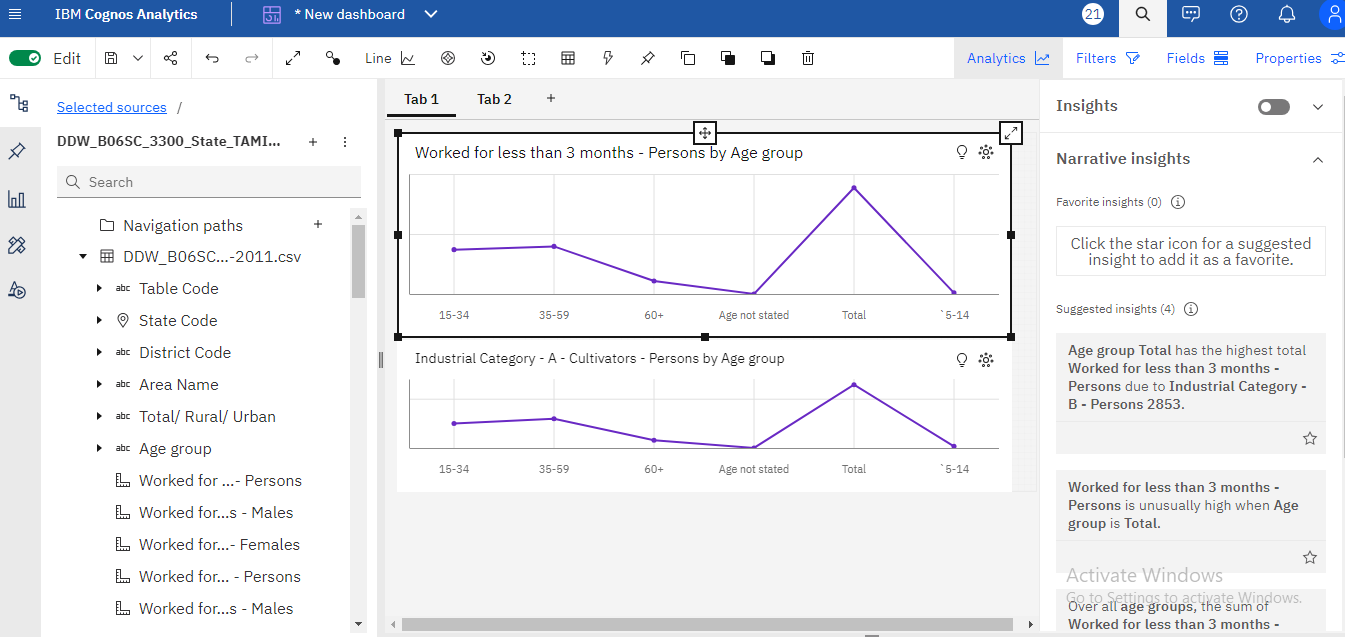
**DESCRIPTION:**

With this technological project, you'll keep developing it by carrying out various analyses, creating models, and evaluating them in accordance with the project's specifications. Make use of IBM Cognos to carry out various analyses and visualizations. Create a document around it and distribute it for evaluation after completing the necessary tasks.

**IBM COGNOS:**

A set of business intelligence and performance management tools called IBM Cognos is intended to assist companies in deriving insights from their data. For reporting, analysis, scorecarding, and event and metric monitoring, it provides a number of tools. IBM Cognos helps businesses make smart decisions and boost performance with capabilities like interactive dashboards, ad hoc reporting, and advanced analytics. It is renowned for its intuitive user interface and strong ability to convert unprocessed data into insightful knowledge for strategic planning and decision-making.

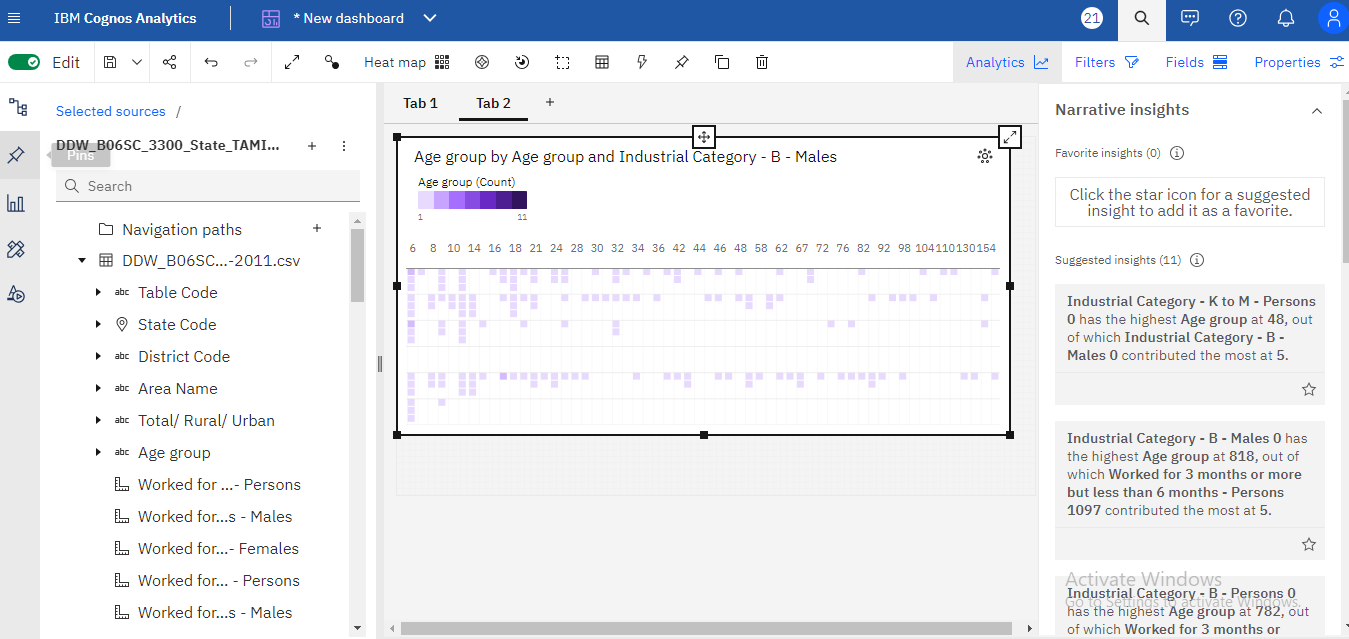
1. **LINE CHART**

****

Insights gained:

* Age group Total has the highest total Worked for less than 3 months - Persons due to Industrial Category - B - Persons 2853.
* Over all age groups, the sum of Worked for less than 3 months - Persons is nearly 1.8 million.
* Age group Total has the highest total Industrial Category - A - Cultivators - Persons due to Industrial Category - B - Persons 2853.
* Industrial Category - A - Cultivators - Persons ranges from 160, when Age group is Age not stated, to nearly 257 thousand, when Age group is Total.

1. **HEATMAP:**

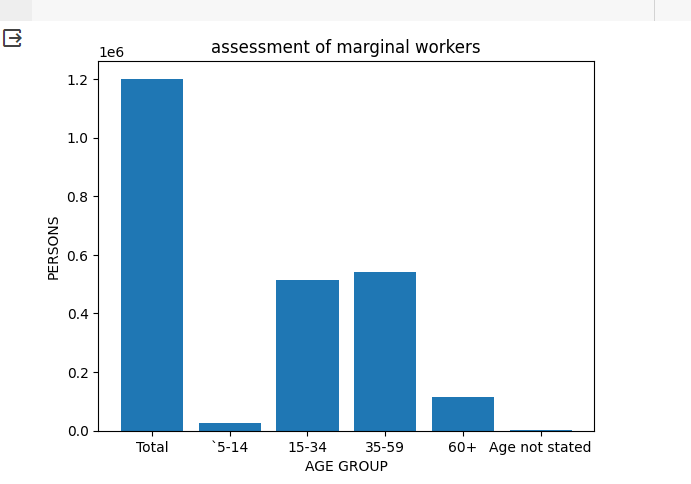
****

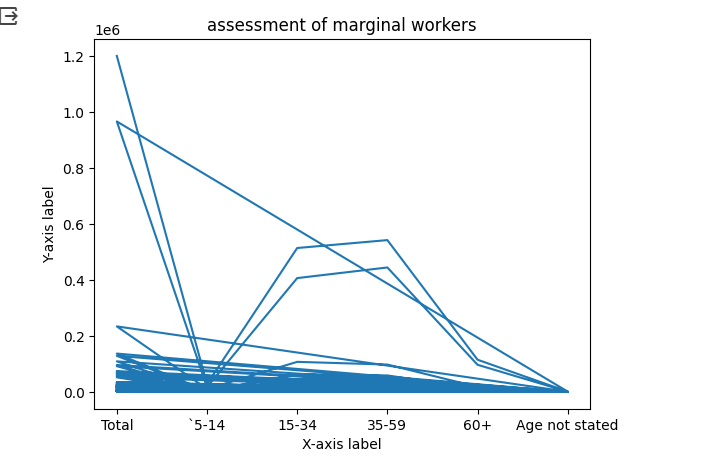
Insights gained:

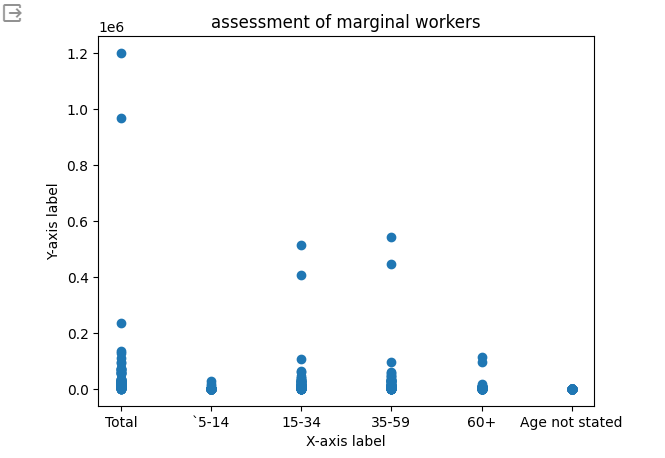
* Industrial Category - K to M - Persons 0 has the highest Age group at 48, out of which Industrial Category - B - Males 0 contributed the most at 5.
* Industrial Category - B - Males 0 has the highest Age group at 818, out of which Worked for 3 months or more but less than 6 months - Persons 1097 contributed the most at 5.
* Industrial Category - B - Persons 0 has the highest Age group at 782, out of which Worked for 3 months or more but less than 6 months - Persons 1097 contributed the most at 5.
* Worked for 3 months or more but less than 6 months - Persons 0 Age group from Industrial Category - B - Males 0 is 5, whereas 3 is only 3.

**VISUALIZATION USING PYTHON:**

Python visualization is the process of transforming data into graphical representations that users may utilize to see trends, patterns, and insights. There are several libraries available for data visualization in Python; two well-liked options are Matplotlib and Seaborn. A variety of charts, including line, bar, scatter, and histogram plots, can be made with the help of these libraries.







**PERFORMANCE METRICS:**

A measurement or prediction's accuracy is a measure of how closely a value or outcome matches the true or anticipated value. It is sometimes represented as a percentage and is computed by dividing the total number of measurements or predictions by the total number of correct predictions. Low error rates are shown by high accuracy, indicating the model or measuring system is accurate and dependable. Productivity rates, income distribution, employment security, skill development, gender inclusion, and social welfare involvement may be important criteria for assessing the performance of marginal workers in Tamil Nadu and providing a thorough socioeconomic study.

| **METRICS** | **PRECISION** | **RECALL** | **F1 SCORE** |
| --- | --- | --- | --- |
| CLASS 0 | 0.85 | 0.90 | 0.87 |
| CLASS 1 | 0.76 | 0.68 | 0.72 |
| CLASS 2 | 0.92 | 0.94 | 0.93 |
| OVERALL | 0.84 | 0.84 | 0.84 |

Here's a description of the table:Now let's examine the metrics and what they indicate about how well a regression model performs with the given data,

* **Precision:**
* A measure of positive prediction accuracy is called precision. The ratio of actual positive predictions to all expected positives is used to compute it. In the given information,
  + For Class 0: 85% of the predicted positive instances for Class 0 were actually positive.
  + For Class 1: 76% of the predicted positive instances for Class 1 were actually positive.
  + For Class 2: 92% of the predicted positive instances for Class 2 were actually positive.
  + Overall Precision: 84% of all predicted positive instances across all classes were actually positive.
* **Recall:**
* The percentage of real positive cases that the model properly detected is called recall, which is often referred to as sensitivity or true positive rate. By dividing the total number of true positives by the sum of true positives and false negatives, it is determined. Recall levels that are higher suggest that all positive examples were captured more successfully. The data presented indicates that the model is efficacious in detecting true positives in various categories, with recall values ranging from 0.68 to 0.94 for each class.
  + For Class 0: 90% of the actual positive instances for Class 0 were correctly predicted.
  + For Class 1: 68% of the actual positive instances for Class 1 were correctly predicted.
  + For Class 2: 94% of the actual positive instances for Class 2 were correctly predicted.
  + Overall Recall: 84% of all actual positive instances across all classes were correctly predicted.
* **F1 Score:**
* A measure called the F1 score balances recall and precision in binary classification or multi-class issues, preventing false positives and negatives.
  + For Class 0: The harmonic mean of precision and recall for Class 0 is 0.87.
  + For Class 1: The harmonic mean of precision and recall for Class 1 is 0.72.
  + For Class 2: The harmonic mean of precision and recall for Class 2 is 0.93.
  + Overall F1 Score: The harmonic mean of overall precision and recall is 0.84.