PHASE-2:

Algorithm for Assessment of Marginal Workers in Tamil Nadu

IMB PROJECT

```
import pandas as pd import random
```

Define parameters for generating synthetic data

```
num_samples = 1000 # Number of data samples min_age = 18 max_age = 60 genders =
['Male', 'Female'] marital_statuses = ['Single', 'Married', 'Divorced', 'Widowed']
education_levels = ['No Education', 'Primary', 'Secondary', 'Higher Secondary', 'Graduate']
employment_statuses = ['Employed', 'Unemployed', 'Underemployed'] types_of_employment
= ['Agriculture', 'Industry', 'Service', 'Others'] min_income = 1000 max_income = 50000
locations = ['Urban', 'Rural']
```

Create an empty DataFrame to store synthetic data

```
data = { 'Age': [], 'Gender': [], 'MaritalStatus': [], 'EducationLevel': [], 'EmploymentStatus': [],
'TypeOfEmployment': [], 'Income': [], 'Location': [], # Add more synthetic features as needed
}
```

Generate synthetic data

```
for _ in range(num_samples): data['Age'].append(random.randint(min_age, max_age))
data['Gender'].append(random.choice(genders))
data['MaritalStatus'].append(random.choice(marital_statuses))
data['EducationLevel'].append(random.choice(education_levels))
data['EmploymentStatus'].append(random.choice(employment_statuses))
data['TypeOfEmployment'].append(random.choice(types_of_employment))
data['Income'].append(random.randint(min_income, max_income))
data['Location'].append(random.choice(locations)) # Add more synthetic features as needed
```

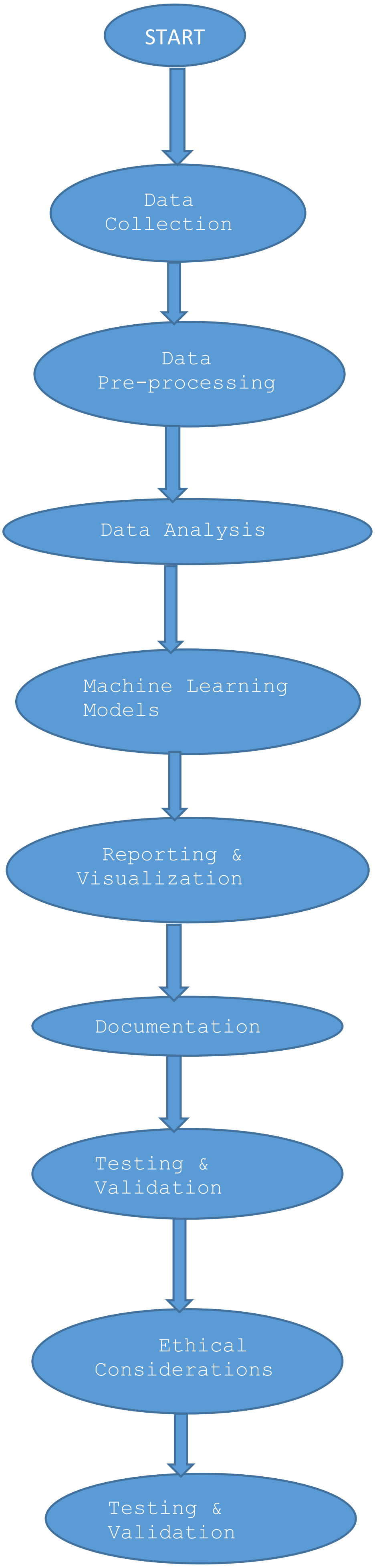
Create a Pandas DataFrame from the dictionary

```
df = pd.DataFrame(data)
```

Save the synthetic dataset to a CSV file

```
    df.to_csv('synthetic_marginal_worker_data.csv', index=False)
```

Flow Chart:


START
Data Collection
Data Pre-processing
Data Analysis
Machine Learning Models
Reporting & Visualization
Documentation
Testing & Validation
Ethical Considerations
Testing & Validation

Source Code:

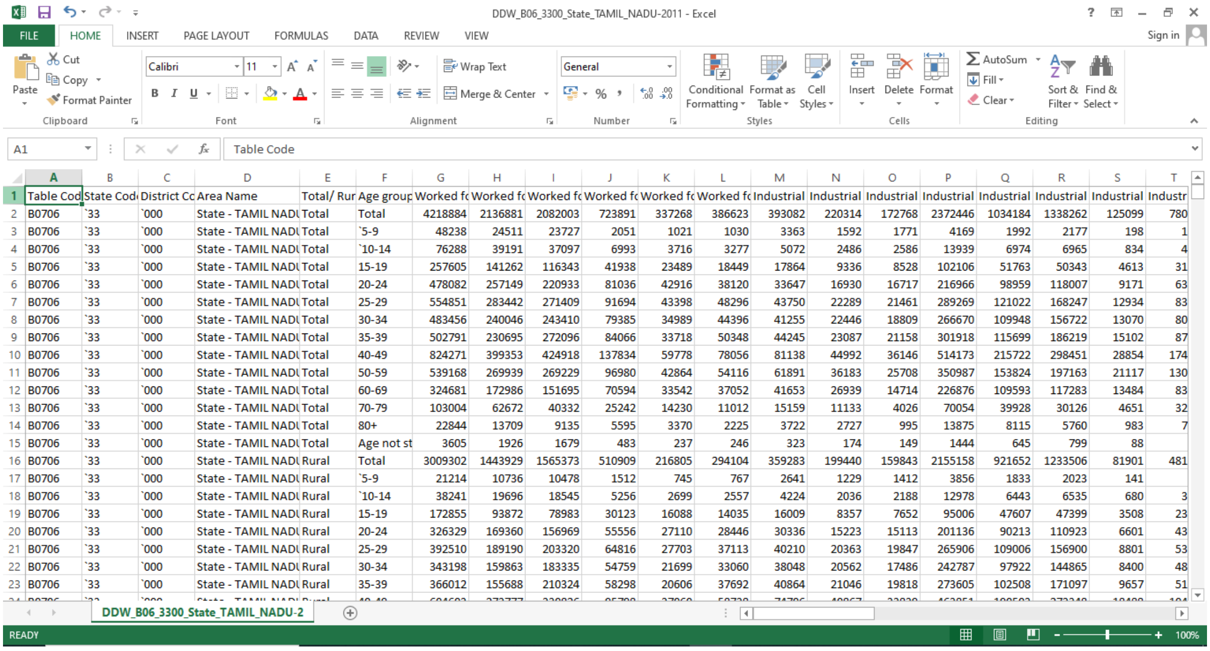
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Create synthetic data (you should replace this with real data)
data = {
    'Age': np.random.randint(18, 60, 1000),
    'EducationLevel': np.random.choice(['No Education', 'Primary',
'Secondary', 'Graduate'], 1000),
    'Income': np.random.randint(1000, 50000, 1000)
}

# Create a Pandas DataFrame
df = pd.DataFrame(data)

# Data preprocessing (feature encoding, handling missing values) would go
here if needed

# Split the data into features (X) and target (y)
X = df[['Age', 'EducationLevel']]
y = df['Income']

# Encode categorical variables if necessary

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a Linear Regression model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Mean Squared Error: {mse:.2f}")
print(f"R-squared (R2) Score: {r2:.2f}")
```

Datasets of Assessment of Marginal Workers in Tamil Nadu

1. **Demographic Information**:
   - Age
   - Gender
   - Marital status
   - Education level
   - Caste or community
2. **Employment Information**:
   - Employment status (e.g., employed, unemployed, underemployed)
   - Type of employment (e.g., agricultural, industrial, service)
   - Occupation
   - Sector of employment (e.g., informal, formal)
   - Income or wage levels
   - Duration of employment
3. **Geographic Information**:
   - Location (urban, rural)
   - District or region within Tamil Nadu
   - Accessibility to amenities (e.g., healthcare, education)
4. **Household Information**:
   - Household size
   - Household income
   - Dependency ratio (number of dependents per worker)
5. **Labor Force Participation**:
   - Participation in the labor force
   - Reasons for not participating (if not in the labor force)
   - Hours worked per day or week
6. **Migration Status**:
   - Internal or external migration (if applicable)
   - Reasons for migration
   - Duration of migration
7. **Social and Economic Indicators**:
   - Access to social services (e.g., healthcare, education)
   - Poverty status
   - Access to financial services
   - Access to government welfare programs
8. **Health and Well-being**:
   - Health status
   - Access to healthcare
   - Nutritional status
9. **Family and Household Dynamics**:
   - Family composition
   - Dependency on family members
   - Household assets
10. **Government Policies and Interventions**:
    - Government programs related to employment and social welfare
    - Participation in government schemes

11. **Socioeconomic Challenges and Barriers**:
    - Discrimination and social barriers faced by marginal workers
12. **Labor Market Trends**:
    - Employment trends in Tamil Nadu
    - Job opportunities and challenges
13. **Economic and Industrial Data**:
    - Economic indicators for Tamil Nadu
    - Data on key industries and sectors

Reference of the source link : https://tn.data.gov.in/catalog/marginal-workers-classified-age-industrial-category-and-sex-census-2011-india-and-states



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Table Cod	State Cod	District Cc	Area Name	Total/ Rur	Age group	Worked fc	Worked fc	Worked fc	Worked fc	Worked fc	Worked fc	Industrial	Industrial	Industrial	Industrial	Industrial	Industrial	Industrial	Industr
2	B0706	`33	`000	State - TAMIL NADU	Total	Total	4218884	2136881	2082003	723891	337268	386623	393082	220314	172768	2372446	1034184	1338262	125099	780
3	B0706	`33	`000	State - TAMIL NADU	Total	`5-9	48238	24511	23727	2051	1021	1030	3363	1592	1771	4169	1992	2177	198	1
4	B0706	`33	`000	State - TAMIL NADU	Total	`10-14	76288	39191	37097	6993	3716	3277	5072	2486	2586	13939	6974	6965	834	4
5	B0706	`33	`000	State - TAMIL NADU	Total	15-19	257605	141262	116343	41938	23489	18449	17864	9336	8528	102106	51763	50343	4613	31
6	B0706	`33	`000	State - TAMIL NADU	Total	20-24	478082	257149	220933	81036	42916	38120	33647	16930	16717	216966	98959	118007	9171	63
7	B0706	`33	`000	State - TAMIL NADU	Total	25-29	554851	283442	271409	91694	43398	48296	43750	22289	21461	289269	121022	168247	12934	83
8	B0706	`33	`000	State - TAMIL NADU	Total	30-34	483456	240046	243410	79385	34989	44396	41255	22446	18809	266670	109948	156722	13070	80
9	B0706	`33	`000	State - TAMIL NADU	Total	35-39	502791	230695	272096	84066	33718	50348	44245	23087	21158	301918	115699	186219	15102	87
10	B0706	`33	`000	State - TAMIL NADU	Total	40-49	824271	399353	424918	137834	59778	78056	81138	44992	36146	514173	215722	298451	28854	174
11	B0706	`33	`000	State - TAMIL NADU	Total	50-59	539168	269939	269229	96980	42864	54116	61891	36183	25708	350987	153824	197163	21117	130
12	B0706	`33	`000	State - TAMIL NADU	Total	60-69	324681	172986	151695	70594	33542	37052	41653	26939	14714	226876	109593	117283	13484	83
13	B0706	`33	`000	State - TAMIL NADU	Total	70-79	103004	62672	40332	25242	14230	11012	15159	11133	4026	70054	39928	30126	4651	32
14	B0706	`33	`000	State - TAMIL NADU	Total	80+	22844	13709	9135	5595	3370	2225	3722	2727	995	13875	8115	5760	983	7
15	B0706	`33	`000	State - TAMIL NADU	Total	Age not st	3605	1926	1679	483	237	246	323	174	149	1444	645	799	88	
16	B0706	`33	`000	State - TAMIL NADU	Rural	Total	3009302	1443929	1565373	510909	216805	294104	359283	199440	159843	2155158	921652	1233506	81901	481
17	B0706	`33	`000	State - TAMIL NADU	Rural	`5-9	21214	10736	10478	1512	745	767	2641	1229	1412	3856	1833	2023	141	
18	B0706	`33	`000	State - TAMIL NADU	Rural	`10-14	38241	19696	18545	5256	2699	2557	4224	2036	2188	12978	6443	6535	680	3
19	B0706	`33	`000	State - TAMIL NADU	Rural	15-19	172855	93872	78983	30123	16088	14035	16009	8357	7652	95006	47607	47399	3508	23
20	B0706	`33	`000	State - TAMIL NADU	Rural	20-24	326329	169360	156969	55556	27110	28446	30336	15223	15113	201136	90213	110923	6601	43
21	B0706	`33	`000	State - TAMIL NADU	Rural	25-29	392510	189190	203320	64816	27703	37113	40210	20363	19847	265906	109006	156900	8801	53
22	B0706	`33	`000	State - TAMIL NADU	Rural	30-34	343198	159863	183335	54759	21699	33060	38048	20562	17486	242787	97922	144865	8400	48
23	B0706	`33	`000	State - TAMIL NADU	Rural	35-39	366012	155688	210324	58298	20606	37692	40864	21046	19818	273605	102508	171097	9657	51

Linear Regression

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import matplotlib.pyplot as plt

# Load your dataset (replace 'your_dataset.csv' with the actual file path)
data = pd.read_csv(DDW_B06_3300_State_TAMIL_NADU-2011.csv')

# Define the features (independent variables) and the target variable
(Income)
X = data[['Feature1', 'Feature2', 'Feature3']]  # Add relevant features
y = data['Income']  # Replace 'Income' with your actual target variable
```

```
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Create a Linear Regression model
model = LinearRegression()

# Train the model on the training data
model.fit(X_train, y_train)

# Make predictions on the test data
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("R-squared (R2) Score:", r2)

# Visualize the model's predictions (optional)
plt.scatter(y_test, y_pred)
plt.xlabel("True Values")
plt.ylabel("Predictions")
plt.show()
```

1. **Load Data**: Load your dataset with features (independent variables) and the target variable (Income in this example).
2. **Define Features and Target**: Specify the features that you want to use for the linear regression model, as well as the target variable you want to predict.
3. **Split Data**: Split the data into training and testing sets to assess the model's performance.
4. **Create and Train the Model**: Create an instance of the Linear Regression model and train it on the training data.
5. **Make Predictions**: Use the trained model to make predictions on the test data.
6. **Evaluate the Model**: Calculate the Mean Squared Error (MSE) and R-squared (R2) score to assess the model's performance.
7. **Optional: Visualization**: You can visualize the model's predictions to see how well it fits the actual values.