***TOPIC***

**Energius : Intelligent Energy Optimization for Greener Tomorrow**

***DATA COLLECTION***

**We shall be working with a synthetic dataset where the idea is to simulate the parameters to mimic real-life scenarios of energy consumption in different households across India. The following are the ways in which we are going to proceed with our objective :**

1. **Defining the parameters :**

**a) Strata ( Categorical Variable )**

**Description : The objective of this parameter is to divide the population into lower, middle and upper classes.**

**b) Demography ( Categorical Variable )**

**Description : The objective of this parameter is to concentrate on densely populated areas of the country. We shall consider 4 main cities : Delhi, Kolkata, Pune and Gujarat.**

**c) Occupancy ( Discrete Variable) purely random**

**Description : Focuses on the no. of individuals in the house.**

**d) No. of Rooms ( Discrete Variable)**

**Criteria :**

* **Poor : 0-2**
* **Lower : 1-3**
* **Middle class: 2-7**
* **Upper class : 4-10**
* **Affluent: 6-15**

**e) Sustainability Preferences ( Categorical Variable)**

**Description : We shall consider 2 categories - ‘Before Knowing Savings Due to Solar Integration’ and ‘After Knowing Savings Due to Solar Integration.**

**f) Income Level : ( Range )**

**Description : Income range will be taken from Government data.**

**Criteria:**

* **<2,00,000- ‘poor’**
* **2,00,000-5,00,000 -> ‘lower class’**
* **5,00,000-20,00,00-> ‘middle class’**
* **20,00,000-9000000-> ‘upper class’**
* **>9000000-> ‘affluent’**

**g) Energy Consumption : ( Continuous)**

**Description : Continuous data on discrete appliances collected from sensor array.**

**h) Energy Sources : (Categorical Variable ) Renewable and Non-renewable sources**

**i) Power Quality : ( Continuous data ) Sensor array**

**j) Environment Condition ( Categorical ) : Weather API**

**k) Cost of Energy : Electricity Bill,and local electricity vendor data(from their websites and slab infos,also look for smart grid or excess electricity resaleability).**

**l) Energy Efficiency ( Categorical ) : Appliance ratings and energy stars.(we will list out all the popular electric appliances with brand and given energy usage and also add a random variable to get the differences in data which may result due to the age of the appliance.**

**m) Appliance Type ( Categorical ) : Names of the appliances in use.**

**Fan,light,washing machine,ac,charger(phone),charger(laptop),washing machine, geyser,microwave,water pump, induction, electric kettles, chimney, water purifier, fridge, TV, Inverter, Miscellaneous, calling bell, desktop**

**Fan power:**

1. **Systhesizing Technique :**

**Step : 1**

**Each demography should be table.strata should be a table.**

**Step : 2**

**We will create a dictionary for each strata that we will be fitting in the model and exhaustively create synthetic data within the average ranges seen in other real time data collection use cases.**

**Step : 3**

**We will create unique random staces for each strata as a standard. For each houses derived from standard we will draw from the random staces of the list of appliances in the strata dictionary. For which each appliances ( electronics ) , we will take the analogue of commercially available data of power usage. Also, we can further increase the non-entropic values by keeping into account the historical power usage in the classified strata as test case.**

**Note : The synthetic dataset that will be created shall be used for model fitting and forecasting charges. And the live sensor data will be used for real time data visualizations.**

**Preferred Simulation Techniques :**

1. **Monte Carlo simulation**
2. **Agent-Based Modelling**
3. **Bootstrapping**
4. **Markov Chain Monte Carlo**
5. **Synthetic Minority Over-Sampling Technique**
6. **Gaussian Mixture Models**
7. **Random walk Simulation**
8. **Latin Hypercube Sampling**

**3) Database Creation :**

**Demography**

**|**

**Strata**

**|**

**Owner details ( Surface level)**

**|**

**Appliances and their consumption**

**We want to create a synthetic dataset where we have the following columns and each column should have 100 observations:**

1. **Owner ID : not a random column. Should be serially continued from 24001-24100**
2. **Demography : Delhi, Kolkata, Gujarat, Pune**
3. **Income level : (ranges should be given in random according to the following)**

**<2,00,000- this range should constitute 2% of the total population**

**2,00,000-5,00,000**

**5,00,000-20,00,00**

**20,00,000-9000000**

**>9000000**

1. **Strata : poor,lower,middle,upper (randomly generated based on income level)**

**<2,00,000- ‘poor’**

**2,00,000-5,00,000 -> ‘lower class’**

**5,00,000-20,00,00-> ‘middle class’**

**20,00,000-9000000-> ‘upper class’**

**>9000000-> ‘affluent’**

1. **No. of rooms: (randomly generated based on strata level according to the following :**

* **0-1 poor**
* **1-3 lower**
* **2-5 middle**
* **4-7 upper**
* **>8 affluent**

**These should be a discrete variable**

1. **No. of Occupants : (randomly generated between 1-10)**
2. **No. of fans : (randomly generated according to strata ) Discrete**

**Poor - 1**

**Lower - 2**

**Middle - (1-4)**

**Upper - (4-10)**

**Affluent - >8**

1. **No. of lights : (randomly generated according to strata ) Discrete**

**Poor - (1-3)**

**Lower - (2-5)**

**Middle - (7-12)**

**Upper - (10-15)**

**Affluent - >15**

**9) No. of ACs : (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower -( 0-1)**

**Middle - (1-3)**

**Upper - (3-7)**

**Affluent - >8**

**10)No. of washing machine: (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - 0**

**Middle - (0-1)**

**Upper - (1-2)**

**Affluent - >2**

**11) No. of induction : (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - 0**

**Middle - (1-2)**

**Upper - (2-5)**

**Affluent - >5**

**12)No. of fridge : (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - (0-1)**

**Middle - (1-2)**

**Upper - (2-4)**

**Affluent - >4**

**13)No. of water pump: (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - (0-1)**

**Middle - (1-2)**

**Upper - (2-4)**

**Affluent - >4**

**14)No. of water purifier : (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - (0-1)**

**Middle - (1-3)**

**Upper - (2-5)**

**Affluent - >5**

**15)No. of geyser : (randomly generated according to strata )discrete**

**Poor - 0**

**Lower - 1**

**Middle - (1-3)**

**Upper - (2-6)**

**Affluent - >8**

**16)No. of TV : (randomly generated according to strata )discrete**

**Poor -( 0-1)**

**Lower - (1-2)**

**Middle - (1-3)**

**Upper - (3-5)**

**Affluent - >5**

**17)No. of Chimney : (randomly generated according to strata )discrete**

**Poor - 0**

**Lower - 0**

**Middle - (1-2)**

**Upper - (1-4)**

**Affluent - >4**

**18)No. of Inverter : (randomly generated according to strata )discrete**

**Poor - 1**

**Lower - 2**

**Middle - (1-4)**

**Upper - (4-10)**

**Affluent - >8**

**19)No. of desktop: (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - 0**

**Middle - (0-2)**

**Upper - (2-5)**

**Affluent - >5**

**20)No. of laptops : (randomly generated according to strata ) discrete**

**Poor - 0**

**Lower - 0**

**Middle - (1-3)**

**Upper - (2-6)**

**Affluent - >6**

**21)No. of mobiles : (randomly generated according to strata ) discrete**

**Poor - (0-1)**

**Lower - (1-3)**

**Middle - (2-6)**

**Upper - (4-8)**

**Affluent - >8**

**22)Miscellaneous : (randomly generated according to strata ) discrete**

**Poor - 0-2**

**Lower - (0-3)**

**Middle - (3-6)**

**Upper - (6-12)**

**Affluent - .>12**