**DATA-PREPROCESSING (ENCODING)**

1. ENCODING :-

Employment status :

Unemployed (not currently looking for work) = 0

Student = 1

Private Sector Employee = 2

Self Employee = 3

Public Sector Employee = 4

Unemployed (currently looking for work) = 5

Gender :

Male = 0

Female = 1

Prefer not to say = 2

Marital Status :

No = 0

Yes = 1

Location (Label encode):

North India = 0

South India = 1

Foreign = 2

For columns on ***personality traits*** - (Notice\_things\_Before, All\_angles\_Before, Sincere\_prod\_Before, Lonely\_Before, Worry\_health\_Before, Charity\_Before, New\_env\_Before, Meeting\_ppl\_Before, Hob\_interests\_Before, Surveys\_Before,Enthu\_Before)

Strongly Disagree = 1

Disagree = 2

Neutral = 3

Agree = 4

Strongly Agree = 5

Other columns on ***spending patterns*** - (Healthy\_Lifestyle\_Before, Save\_all\_money\_Before , Active\_Investor\_Before, Edu\_instit\_fee\_Before, Spend\_on\_gadgets\_Before)

Strongly Disagree = 1

Disagree = 2

Neutral = 3

Agree = 4

Strongly Agree = 5

For columns on ***Interests*** - (Pol\_interest\_Before, Internet\_interest\_Before , Economy\_Manag\_intrst\_Before, Medicine\_intrst\_Before , Religion\_intrst\_Before)

Not interested = 1

Neutral = 2

Moderately Interested = 3

Very Interested = 4

***Miscellaneous*** ***1***- (WFH\_office\_Before, Brand\_non-brand\_Before, Basic\_medications\_Before, Digital\_content\_Before, spend\_on\_Luxury\_Before)

Disagree = 2

Neutral = 3

Agree = 4

Spending\_kind\_Before :

Low spending = 0

Medium spending = 1

High spending = 2

Track\_Household\_exp\_Before :

No = 0

Maybe = 1

Yes = 2

Mode\_of\_transport\_Before :

Public (Metro, govt e.t.c) = 0

Private (cabs e.t.c) = 1

Personal = 2

Food\_pref\_Before :

Outside (Online & Offline) = 0

Home - Cooked = 1

Medi\_pref\_Before :

Homemade or Traditional (Ayurvedic e.t.c) = 0

English Medicine = 1

Smoking\_hab\_Before :

Never Smoked = 0

Former Smoker = 1

Tried smoking = 2

Current smoker = 3

Drinking\_hab\_Before :

Never = 0

Social Drinker = 1

Drink a lot = 2

Sleeping\_hab\_Before :

Early to bed and early to rise = 0

Late to bed and early to rise = 1

Early to bed and late to rise = 2

late to bed and late to rise = 3

Spent\_onli\_Before :

No time at all = 0

Less than an hour a day = 1

Few hours a day = 2

Most of the day = 3

Income\_Change, Change\_in\_data\_consumption :

Decreased = 0

Same = 1

Increased = 2

* + - Age :

20-25 = 0

25-30 = 1

30-35 = 2

35-40 = 3

40-55 = 4

55+ = 5

* + - Income :

5 - 10 lakh per annum = 3,

3 - 5 lakh per annum = 2,

Under 3 lakh per annum = 1,

I don't earn right now (maybe in future) = 0,

10 - 20 lakh per annum = 4,

20+ lakh per annum = 5