Data set of Cancer, Population, and Cigadetails

Cancer Dataset

- This dataset collected from the website online lung cancer predication system
- Total no. of attribtes:16, No. of instances: 284

  - Gender: M(male), F(female0)
  - Age: Age of the patient
  - Smoking: YES=2 , NO=1.
  - Yellow fingers: YES=2 , NO=1.
  - Anxiety: YES=2 , NO=1.
  - Peer_pressure: YES=2 , NO=1.
  - Chronic Disease: YES=2 , NO=1.
  - Fatigue: YES=2 , NO=1.
  - Allergy: YES=2 , NO=1.
  - Wheezing: YES=2 , NO=1.
  - Alcohol: YES=2 , NO=1.
  - Coughing: YES=2 , NO=1.
  - Shortness of Breath: YES=2 , NO=1.
  - Swallowing Difficulty: YES=2 , NO=1.
  - Chest pain: YES=2 , NO=1.
  - Lung Cancer: YES , NO.

- This dataset contains information on patients with lung cancer, including their age, gender, air pollution exposure, alcohol use, dust allergy, occupational hazards, genetic risk, chronic lung disease, balanced diet, obesity, smoking, passive smoker, chest pain, coughing of blood, fatigue, weight loss ,shortness of breath ,wheezing ,swallowing difficulty ,clubbing of finger nails and snoring.
- Total no. of attributes: 26 , No. of instances: 1000

  - Index
  - Patient Id
  - Age: The age of the patient. (Numeric)
  - Gender: The gender of the patient. (Categorical)
  - Air Pollution: The level of air pollution exposure of the patient. (Categorical)
  - Alcohol use: The level of alcohol use of the patient. (Categorical)
  - Dust Allergy: The level of dust allergy of the patient. (Categorical)
  - OccuPational Hazards: The level of occupational hazards of the patient. (Categorical)
  - Genetic Risk: The level of genetic risk of the patient. (Categorical)
  - chronic Lung Disease: The level of chronic lung disease of the patient. (Categorical)
  - Balanced Diet: The level of balanced diet of the patient. (Categorical)
  - Obesity: The level of obesity of the patient. (Categorical)
  - Smoking: The level of smoking of the patient. (Categorical)
  - Passive Smoker: The level of passive smoker of the patient. (Categorical)
  - Chest Pain: The level of chest pain of the patient. (Categorical)

- Coughing of Blood: The level of coughing of blood of the patient. (Categorical)
- Fatigue: The level of fatigue of the patient. (Categorical)
- Weight Loss: The level of weight loss of the patient. (Categorical)
- Shortness of Breath: The level of shortness of breath of the patient. (Categorical)
- Wheezing: The level of wheezing of the patient. (Categorical)
- Swallowing Difficulty: The level of swallowing difficulty of the patient. (Categorical)
- Clubbing of Finger Nails: The level of clubbing of finger nails of the patient. (Categorical)

- This dataset consists of 100 observations and 10 variables(out of which 8 numeric variables and one categorical variable and ID)
- Total no. of attributes: 10 , No. of instances:100

- id
- diagnosis_result
- radius
- texture
- perimeter
- area
- smoothness
- compactness
- symmetry
- fractal_dimension

Population Dataset

- This dataset is the India Population Density 2011.
- Total no. of attributes: 12, No. of instances: 36

- Rank: 36
- State/UT: 36
- Population
- Percent
- Male
- Female
- Difference Between
- Sex Ratio
- Rural
- Urban
- Area(km2)
- Density(per km2)

-This dataset contains the information of every nation's population from 1960 to 2021

- Total no. of attributes: 64, No. of instances: 266

- Country Name

- Country COde
- 1960 population
- 1961
- …
- 2020
- 2021

- This dataset contains the information of various contry's with more growth rate, area.
- It contain 3 csv file with named
  `2021_population.csv`,`2022_population.csv`,`2023_population.csv`.
- Total no. of attributes: 9 , No of instances: 223

- ISO Code
- Country
- 2021_last_update (last updated data for 2021)
- 2020_population (Population in 2020)
- area (Area(square KM))
- density_sq_km (Density per squaer KM)
- growth_rate
- world_% (World Population percentage)
- rank

Cigarettes Dataset

- This data set contains 1125 observations of 9 variables associated with cigarettes sold between 1998 and 2000. The idea for creating this data set came from this paper by Lauren McIntyre, where she analyzed the carbon monoxide output of 25 cigarettes from different brands. features:

- Brand_Name : Brand of cigarette manufacture.
- Flavor : Description of the cigarette name such as *Gold*, *Plain* or *Slim*. 43 distinct values
- CO : mg of CO, rounded to nearest mg
- nic : mg of nicotine, rounded to nearest tenth of mg
- tar : mg of tar, rounded to nearest mg
- size : descriptor of cigarette, one of 100, King, Reg, 120, or 80
- filter : Does the cigarette have a filter?
- pack : hard pack, soft pack
- mentho : lis the cigarette menthol?