**/Library/Frameworks/Python.framework/Versions/3.11/lib/python3.11/site-packages/aif360/data/raw/german**

**https://archive.ics.uci.edu/datasets?search=Heart%20Disease**

**https://archive.ics.uci.edu/dataset/503/hepatitis+c+virus+hcv+for+egyptian+patients**

**Heart Attack Analysis & Prediction Dataset**

<https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

Samples: 303

Attributes: 10

* Age: Age of the patient
* **Sex: Sex of the patient (0 = female; 1 = male) (female = 96 samples; male = 207 samples)**
* exang: exercise induced angina (1 = yes; 0 = no) (yes = 99 samples; no = 204 samples)
* ca: number of major vessels (0-3)
  + Value 0 (175 samples)
  + Value 1 (65 samples)
  + Value 2 (38 samples)
  + Value 3 (20 samples)
  + Value 4 (5 samples)
* cp: chest pain type
  + Value 0: typical angina (143 samples)
  + Value 1: atypical angina (50 samples)
  + Value 2: non-anginal pain (87 samples)
  + Value 3: asymptomatic (23 samples)
* trtbps: resting blood pressure (in mm Hg)
* chol: cholestoral in mg/dl fetched via BMI sensor
* fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false) (1 = 45 samples; 0 = 258 samples)
* rest\_ecg: resting electrocardiographic results
  + Value 0: normal (147 samples)
  + Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) (152 samples)
  + Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria (4 samples)
* thalach: maximum heart rate achieved
* **target: 0= less chance of heart attack (138 samples); 1= more chance of heart attack (165 samples)**

**Airline Passenger Satisfaction**

<https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

Samples: 25976

Attributes: 22 (there are 24, but the first two must be removed because they are identifying codes)

* Gender: Gender of the passengers (Female = 51%; Male = 49%)
* **Customer Type: The customer type (Loyal customer = 82%; disloyal customer = 18%)**
* Age: The actual age of the passengers
* **Type of Travel: Purpose of the flight of the passengers (Personal Travel = 31%; Business Travel = 69%)**
* Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)
* Flight distance: The flight distance of this journey
* Inflight wifi service: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
* Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient
* Ease of Online booking: Satisfaction level of online booking
* Gate location: Satisfaction level of Gate location
* Food and drink: Satisfaction level of Food and drink
* Online boarding: Satisfaction level of online boarding
* Seat comfort: Satisfaction level of Seat comfort
* Inflight entertainment: Satisfaction level of inflight entertainment
* On-board service: Satisfaction level of On-board service
* Leg room service: Satisfaction level of Leg room service
* Baggage handling: Satisfaction level of baggage handling
* Check-in service: Satisfaction level of Check-in service
* Inflight service: Satisfaction level of inflight service
* Cleanliness: Satisfaction level of Cleanliness
* Departure Delay in Minutes: Minutes delayed when departure
* Arrival Delay in Minutes: Minutes delayed when Arrival
* **Satisfaction (target): Airline satisfaction level (Satisfied = 44%; neutral or dissatisfied = 56%)**

**Students Performance**

<https://www.kaggle.com/datasets/joebeachcapital/students-performance>

Samples: 145

Attributes: 31

* Student Age (1: 18-21, 65 samples; 2: 22-25, 70 samples; 3: above 26, 10 samples)
* **Sex (1: female, 58 samples; 2: male, 87 samples)**
* Graduated high-school type: (1: private, 25 samples; 2: state, 103 samples; 3: other, 17 samples)
* Scholarship type: (1: None, 2: 25%, 3: 50%, 4: 75%, 5: Full)
* Additional work: (1: Yes, 49 samples; 2: No, 96 samples)
* Regular artistic or sports activity: (1: Yes, 58 samples; 2: No, 87 samples)
* Do you have a partner: (1: Yes, 61 samples; 2: No, 84 samples)
* Total salary if available (1: USD 135-200, 2: USD 201-270, 3: USD 271-340, 4: USD 341-410, 5: above 410)
* Transportation to the university: (1: Bus, 2: Private car/taxi, 3: bicycle, 4: Other)
* Accommodation type in Cyprus: (1: rental, 2: dormitory, 3: with family, 4: Other)
* Mothers education: (1: primary school, 2: secondary school, 3: high school, 4: university, 5: MSc., 6: Ph.D.)
* Fathers education: (1: primary school, 2: secondary school, 3: high school, 4: university, 5: MSc., 6: Ph.D.)
* Number of sisters/brothers (if available): (1: 1, 2:, 2, 3: 3, 4: 4, 5: 5 or above)
* Parental status: (1: married, 2: divorced, 3: died - one of them or both)
* Mothers occupation: (1: retired, 2: housewife, 3: government officer, 4: private sector employee, 5: self-employment, 6: other)
* Fathers occupation: (1: retired, 2: government officer, 3: private sector employee, 4: self-employment, 5: other)
* Weekly study hours: (1: None, 2: <5 hours, 3: 6-10 hours, 4: 11-20 hours, 5: more than 20 hours)
* Reading frequency (non-scientific books/journals): (1: None, 2: Sometimes, 3: Often)
* Reading frequency (scientific books/journals): (1: None, 2: Sometimes, 3: Often)
* Attendance to the seminars/conferences related to the department: (1: Yes, 2: No)
* Impact of your projects/activities on your success: (1: positive, 2: negative, 3: neutral)
* Attendance to classes (1: always, 2: sometimes, 3: never)
* Preparation to midterm exams 1: (1: alone, 2: with friends, 3: not applicable)
* Preparation to midterm exams 2: (1: closest date to the exam, 2: regularly during the semester, 3: never)
* Taking notes in classes: (1: never, 2: sometimes, 3: always)
* Listening in classes: (1: never, 2: sometimes, 3: always)
* Discussion improves my interest and success in the course: (1: never, 2: sometimes, 3: always)
* Flip-classroom: (1: not useful, 2: useful, 3: not applicable)
* Cumulative grade point average in the last semester (/4.00): (1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49)
* Expected Cumulative grade point average in the graduation (/4.00): (1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49)
* Course ID
* **Grade (target): 0: Fail, 1: DD, 2: DC, 3: CC, 4: CB, 5: BB, 6: BA, 7: AA**

**Loan Data**

<https://www.kaggle.com/datasets/zhijinzhai/loandata>

Samples: 500

Attributes: 31

* Loan\_id: A unique loan number assigned to each loan customers
* **Loan\_status (target): Whether a loan is paid off (60%), in collection (20%), new customer yet to payoff or paid off after the collection efforts (20%)**
* Principal: Basic principal loan amount at the origination
* Terms: Can be weekly (7 days), biweekly, and monthly payoff schedule
* Effective\_date: When the loan got originated and took effects
* Due\_date: Since it’s one-time payoff schedule, each loan has one single due date
* Paidoff\_time: The actual time a customer pays off the loan
* Pastdue\_days: How many days a loan has been past due
* Age:
* Education: College (44%); High School or Below (42%); Other (14%)
* **Gender: Male (85%); Female (15%)**

**Hepatitis C Prediction Dataset**

<https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset>

Samples: 615

Attributes: 14

* X (Patient ID/No.)
* **Category (diagnosis, target): '0 = Blood Donor', 87%; '0s = suspect Blood Donor'; '1 = Hepatitis'; '2 = Fibrosis'; '3 = Cirrhosis', 5%)**
* Age (in years)
* **Sex (f = 39%; m = 61%)**
* Attributes 5 to 14 refer to laboratory data:
  + ALB
  + ALP
  + ALT
  + AST
  + BIL
  + CHE
  + CHOL
  + CREA
  + GGT
  + PROT