

Dataset Descriptions and Data Dictionaries

1 Climate dataset

Filename: climate.csv

Description: The dataset contains climate or atmospheric data, and can be used to investigate the impact of various factors on the average global temperature. Each row corresponds to climate or atmospheric data captured for a given month. The dataset contains 308 rows and 11 columns.

Name	Description	Type	Details
Year	Year of observation	Categorical	
Month	Month of observation	Categorical	
MEI	Multivariate El Nino Southern Oscillation index (MEI)	Numerical	A measure of the strength of the El Nino/La Nina-Southern Oscillation
CO2	Atmospheric concentration of carbon dioxide	Numerical	Measured in parts per million by volume
CH4	Atmospheric concentration of methane	Numerical	Measured in parts per million by volume
N2O	Level of atmospheric concentration of nitrous oxide	Categorical	Possible values: "H", "L"
CFC-11	Atmospheric concentration of trichlorofluoromethane	Numerical	Measured in parts per billion by volume
CFC-12	Atmospheric concentration of dichlorodifluoromethane	Numerical	Measured in parts per billion by volume
TSI	Level of Total Solar Irradiance. TSI is the rate at which sun's energy is deposited per unit area	Categorical	Possible values: "H", "M", "L"
Aerosols	Level of stratospheric "aerosol optical depth" at 550nm. This is linked to volcanoes, as eruptions result in new particles added to atmosphere	Categorical	Possible values: "H", "M", "L"
Temp	Difference in temperature between average global temperature and a reference value	Numerical	

2 Credit dataset

Filename: credit.csv

Description: The dataset contains data of individuals applying for a credit facility in a bank. Each row corresponds to a different applicant and contains attributes captured by bank during an application. The dataset contains 1000 rows and 20 columns.

Name	Description	Type	Details
ID	Identification number	Categorical	
Checking_Account	Status of checking account	Categorical	1: \$0; 2: <\$200; 3: >=\$200 4: no checking account
Duration	Credit duration in months	Numerical	
Payment_Status	Credit history	Categorical	0: no credits taken 1: all credits paid back 2: existing credits paid back till now 3: delay in paying off in the past 4: critical account
Purpose	Purpose of credit	Categorical	0: car (new); 1: car (used) 2: furniture/equipment; 3: radio/television 4: domestic appliances; 5: repairs 6: education; 7: vacation; 8: retraining 9: business; 10: others
Amount	Credit amount	Numerical	
Savings_Account	Status of savings account	Categorical	1: < \$100; 2: \$100 - \$500 3: \$500 - \$1000; 4: >\$3000 5: unknown/no savings account
Employment	Length of current employment	Categorical	1: unemployed; 2: <1 year 3: 1 - 4 years; 4: 4 - 7 years; 5: >7 years
Installment	Installment rate as percentage of disposable income	Numerical	
Personal_Status	Marital status and sex	Categorical	1: male, divorced/separated 2: male, single 3: male, married/widowed 4: female
Guarantors	Other debtors or guarantors	Categorical	1: none; 2: co-applicant; 3: guarantor
Residence_Length	Number of years staying in current residence	Categorical	1: < 1 year; 2: 1-2 years; 3: 2-3 years 4: >3 years
Assets	Asset ownership	Categorical	1: real estate 2: savings agreement/life insurance (if no 1) 3: car or other (if no 1, 2); 4: unknown/none
Age	Age in years	Numerical	
Credit_Rating	Credit rating	Numerical	
Existing_Credits	Other existing credit in place	Categorical	1: bank; 2: stores; 3: none
Housing_Type	Type of apartment	Categorical	1: rent; 2: own; 3: for free
Num_Credits	Number of existing credits	Numerical	
Occupation	Occupation	Categorical	1 : unemployed/ unskilled - non-resident 2 : unskilled – resident 3 : skilled employee / official 4 : management/ self-employed
Dependents	Number of dependents	Numerical	
Telephone	Has telephone	Categorical	1: none; 2: yes, registered to customer
Foreign_Worker	Foreign worker or not	Categorical	1: yes; 2: no
Approval	Loan approval status	Categorical	1: approved; 0: not approved

3 Employee dataset

Filename: employee.csv

Description: The dataset contains HR data related to a sales team. Specifically the dataset captures employees' information at the beginning of each month, including if the employee quit in that given month, total sales acquired so far in his/her career and his/her latest quarterly rating. The dataset contains 2381 rows and 13 columns.

Name	Description	Type	Details
Date	Record capture date	Categorical	Date of format dd/mm/yyyy
Emp_ID	Employee ID	Categorical	
Age	Age	Numerical	
Gender	Gender	Categorical	"Male", "Female"
City	City where employee works	Categorical	"C1", "C2", ..., "C29"
Education	Highest education level of employee	Categorical	"College", "Bachelor", "Master"
Salary	Last drawn salary	Numerical	
Join_Date	Date of joining the sales team	Categorical	Date of format dd/mm/yyyy
Last_Work_Date	Last working date with the sales team	Categorical	Date of format dd/mm/yyyy. If employee does not quit in the given month where the record is captured, this field would be empty
Join_Designation	Designation level when first joined	Categorical	1, 2, 3, 4, 5
Designation	Designation level at the time record is captured	Categorical	1, 2, 3, 4, 5
Total_Sales_Acquired	Total sales generated by employee since joining the team	Numerical	
Quarterly_Rating	Latest quarterly performance rating	Categorical	1, 2, 3, 4