CS105 Statistical Thinking for Data Science, 2022-23 Term 2 School of Computing and Information Systems, Singapore Management University

## **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<br>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.<br>This is linked to volcanoes, as eruptions result in new<br>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 23 columns.

| Name             | Description                | Type        | Details   |
|------------------|----------------------------|-------------|---|
| ID               | Identification number      | Categorical |   |
| Checking_Account | Status of checking account | Categorical | 1: \$0; 2: <\$200; 3: >=\$200<br>4: no checking account |
| Duration         | Credit duration in months  | Numerical   |   |

CS105 Statistical Thinking for Data Science, 2022-23 Term 2 School of Computing and Information Systems, Singapore Management University

| Purpose   Purpose of credit   Categorical   Categorical   Categorical   Categorical   Purpose   Purpose of credit   Categorical   Categorica   | Payment_Status   | Credit history                                      | Categorical | 0: no credits taken  |
|--|------------------|---|-------------|--|
| Purpose   Purpose of credit   Purpose of credit   Purpose of credit   Purpose   Purpose   Purpose of credit   Purpose   Purp   | 1 uj mem_s tatas |   | Curegorieur |  |
| Purpose Purpose of credit Categorical Purpose of credit Categorical Purpose Purpose of credit Categorical Categori |                  |   |             |  |
| Purpose Purpose of credit  |                  |   |             |  |
| Purpose       Purpose of credit       Categorical 2 (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 3: \$\$100; 2: \$100; 5500 3: \$500 \$\$1000; 4: \$\$3000 5: unknown/no savings account         Employment       Length of current employment       Categorical 1: unemployed; 2: 4! year 3: 1 - 4 years; 4: 4 - 7 years; 5: >7 year         Installment       Installment rate as percentage of disposable income       Numerical         Personal_Status       Marital status and sex       Categorical 2: 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: <2 years; 3: 2-3 years 4: >3 years         Age       Age in years       Numerical 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown 4: >3 years 4: >3 years 4: >3 years 5: >3 years 7: >3 years 7  |                  |   |             |  |
| 2: furniture/equipment; 3: radio/televis 4: domestic appliances; 5: repairs 6: education; 7: vacation; 8: retraining 9: business; 10: others    Amount   Credit amount   Numerical   | Purpose          | Purpose of credit                                   | Categorical |  |
| 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 1: unemployed; 2: <1 year 3: 1 - 4 years; 4: 4 - 7 years; 5: >7 year 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: real estate 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown Age Age in years  Credit_Rating Credits Other existing credit in place Categorical 1: rent; 2: own; 3: for free  | r urpose         | r dipose of credit                                  | Categorical |  |
| 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 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: real estate 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown Age Age in years  Credit_Rating Credits Other existing credit in place Categorical 1: rent; 2: own; 3: for free   |                  |   |             |  |
| Amount Credit amount Numerical  Savings_Account Status of savings account Categorical 1: < \$100; 2: \$100 - \$500 3: \$500 - \$1000; 4: >\$3000 } 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 year  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 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical Credit rating Numerical Value between 1 and 100  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  |                  |   |             |  |
| Amount       Credit amount       Numerical         Savings_Account       Categorical       1: <\$100; 2: \$100 - \$500   |                  |   |             |  |
| Savings_Account  Status of savings account  Categorical  Employment  Length of current employment  Installment  Installment rate as percentage of disposable income  Personal_Status  Marital status and sex  Categorical  Categorical  I: unemployed; 2: <1 year 3: 1 - 4 years; 4: 4 - 7 years; 5: >7 year  Installment  Personal_Status  Marital status and sex  Categorical  Categorical  I: male, divorced/separated 2: male, single 3: male, married/widowed 4: female  Guarantors  Other debtors or guarantors  Categorical  Categorical  I: <1 year; 2: 1-2 years; 3: 2-3 years 4: >3 years  Assets  Asset ownership  Categorical  Categorical  I: real estate 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown  Age  Age in years  Numerical  Credit_Rating  Credit rating  Numerical  Numerical  Value between 1 and 100  Existing_Credits  Other existing credit in place  Categorical  I: rent; 2: own; 3: for free  | Amount           | Credit amount                                       | Numerical   | 7. business, 10. outers  |
| Semployment   Length of current employment   Categorical   1: unemployed; 2: <1 year   3: 1 - 4 years; 4: 4 - 7 years; 5: >7 year  | Amount           | Credit amount                                       | rumencai    |  |
| Employment Length of current employment Categorical 1: unemployed; 2: <1 year 3: 1 - 4 years; 4: 4 - 7 years; 5: >7 year 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 3: car or other (if no 1, 2); 4: unknown Age Age in years Numerical Credit_Rating Credit rating Numerical Value between 1 and 100  Existing_Credits Other existing credit in place Categorical 1: rent; 2: own; 3: for free   | Savings_Account  | Status of savings account                           | Categorical |  |
| EmploymentLength of current employmentCategorical<br>3: 1 - 4 years; 4: 4 - 7 years; 5: >7 yearInstallmentInstallment rate as percentage of disposable incomeNumericalPersonal_StatusMarital status and sexCategorical<br>2: male, single<br>3: male, married/widowed<br>4: female1: male, divorced/separated<br>2: male, single<br>3: male, married/widowed<br>4: femaleGuarantorsOther debtors or guarantorsCategorical1: none; 2: co-applicant; 3: guarantorResidence_LengthNumber of years staying in current residenceCategorical1: < 1 year; 2: 1-2 years; 3: 2-3 years<br>4: >3 yearsAssetsAsset ownershipCategorical1: real estate<br>2: savings agreement/life insurance (if<br>3: car or other (if no 1, 2); 4: unknownAgeAge in yearsNumericalValue between 1 and 100Credit_RatingCredit ratingNumericalValue between 1 and 100Existing_CreditsOther existing credit in placeCategorical1: bank; 2: stores; 3: noneHousing_TypeType of apartmentCategorical1: rent; 2: own; 3: for free   |                  |   |             | 3: \$500 - \$1000; 4: >\$3000  |
| 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 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical  Credit_Rating Credit rating Numerical Value between 1 and 100  Existing_Credits Other existing credit in place Categorical 1: rent; 2: own; 3: for free   |                  |   |             | 5: unknown/no savings account  |
| 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 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical  Credit_Rating Credit rating Numerical Value between 1 and 100  Existing_Credits Other existing credit in place Categorical 1: rent; 2: own; 3: for free   | Employment       | Length of current employment                        | Categorical | 1: unemployed; 2: <1 year  |
| 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 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical Value between 1 and 100  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  |                  |   |             | 3: 1 - 4 years; 4: 4 - 7 years; 5: >7 years  |
| 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 3: car or other (if no 1, 2); 4: unknown  Age  Age in years  Numerical  Credit_Rating  Credit rating  Numerical  Value between 1 and 100  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  | Installment      | Installment rate as percentage of disposable income | Numerical   | , and the second |
| 2: male, single 3: male, married/widowed 4: female  Guarantors  Other debtors or guarantors  Categorical  Categorical  1: < 1 year; 2: 1-2 years; 3: 2-3 years 4: >3 years  Assets  Asset ownership  Categorical  Categorical  1: < 1 year; 2: 1-2 years; 3: 2-3 years 4: >3 years  Categorical  1: real estate 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown  Age  Age in years  Numerical  Credit_Rating  Credit rating  Numerical  Value between 1 and 100  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   | Darsonal Status  | Marital status and say                              | Catagorical | 1: male_divorced/caparated   |
| Guarantors Other debtors or guarantors Categorical   | i eisonai_Status | Wartar status and sex                               | Categoricai |  |
| Guarantors Other debtors or guarantors Categorical Cat |                  |   |             |  |
| GuarantorsOther debtors or guarantorsCategorical1: none; 2: co-applicant; 3: guarantorResidence_LengthNumber of years staying in current residenceCategorical1: < 1 year; 2: 1-2 years; 3: 2-3 years   |                  |   |             |  |
| 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 3: car or other (if no 1, 2); 4: unknown  Age Age in years  Credit_Rating Credit rating  Numerical Value between 1 and 100  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   | _                |   |             |  |
| Assets Asset ownership Categorical Asset ownership Categorical 1: real estate 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical Credit_Rating Credit rating Numerical Value between 1 and 100  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   | Guarantors       | Other debtors or guarantors                         | Categorical | 1: none; 2: co-applicant; 3: guarantor   |
| Assets Asset ownership Categorical Asset ownership Categorical Categorical Categorical Categorical Categorical Categorical Categorical Categorical Credit_Rating Credit_Rating Credit rating Credit_rating Credit rating Credit rating Credit rating Credit rating Credit rating Categorical Categoric | Residence Length | Number of years staying in current residence        | Categorical | 1: < 1 year; 2: 1-2 years; 3: 2-3 years  |
| Assets Asset ownership Categorical 1: real estate 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical Credit_Rating Credit rating Numerical Value between 1 and 100  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   |                  |   | 3           |  |
| 2: savings agreement/life insurance (if 3: car or other (if no 1, 2); 4: unknown  Age Age in years Numerical  Credit_Rating Credit rating Numerical Value between 1 and 100  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  | Assets           | Asset ownership                                     | Categorical |  |
| Age Age in years  Credit_Rating  Credit rating  Numerical  Numerical  Value between 1 and 100  Existing_Credits  Other existing credit in place  Housing_Type  Type of apartment  Categorical  1: rent; 2: own; 3: for free  |                  | <b>F</b>  | 2 8         |  |
| AgeAge in yearsNumericalCredit_RatingCredit ratingNumericalValue between 1 and 100Existing_CreditsOther existing credit in placeCategorical1: bank; 2: stores; 3: noneHousing_TypeType of apartmentCategorical1: rent; 2: own; 3: for free   |                  |   |             |  |
| Credit_Rating Credit rating Numerical Value between 1 and 100  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  | Age              | Age in years  | Numerical   | or our or outer (if no 1, 2), it diffuse with none   |
| 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   | 1150             | Tigo in yours                                       | rumeneur    |  |
| Housing_Type Type of apartment Categorical 1: rent; 2: own; 3: for free  | Credit_Rating    | Credit rating                                       | Numerical   | Value between 1 and 100  |
| Housing_Type Type of apartment Categorical 1: rent; 2: own; 3: for free  | Existing Credits | Other existing credit in place                      | Categorical | 1: hank: 2: stores: 3: none  |
|  | Laisting_Credits | Other existing eredit in place                      | Categorical | 1. bank, 2. stores, 3. none  |
|  | Housing_Type     | Type of apartment                                   | Categorical | 1: rent; 2: own; 3: for free   |
|  |                  |   |             |  |
| Num_Credits  | Num_Credits      | Number of existing credits                          | Numerical   |  |
| Occupation Occupation Categorical 1 : unemployed/ unskilled - non-reside   | Occupation       | Occupation  | Categorical | 1 : unemployed/ unskilled - non-resident   |
| 2 : unskilled – resident   | Occupation       | Occupation  | Categorical |  |
| 3 : skilled employee / official  |                  |   |             |  |
|  |                  |   |             |  |
| 4 : management/ self-employed  | D 1 /            | N. I. C.I. I.                                       | NT ' 1      | 4: management/ self-employed   |
| Dependents Number of dependents Numerical  | Dependents       | Number of dependents                                | Numerical   |  |
| Telephone Has telephone Categorical 1: none; 2: yes, registered to customer  | Telephone        | Has telephone                                       | Categorical | 1: none; 2: yes, registered to customer  |
|  |                  |   |             |  |
| Foreign_Worker Foreign worker or not Categorical 1: yes; 2: no   | Foreign_Worker   | Foreign worker or not                               | Categorical | 1: yes; 2: no  |
| Approval Loan approval status Categorical 1: approved; 0: not approved   | Approval         | Loan approval status                                | Categorical | 1: approved: 0: not approved   |
| Tapprovai Loan approvai status Categoricai 1. approved, v. not approved  | rippiovai        | Loui approvai status                                | Categorical | 1. αρριόνου, ο. ποι αρριόνου   |

CS105 Statistical Thinking for Data Science, 2022-23 Term 2 School of Computing and Information Systems, Singapore Management University

## 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   |   |
| Ouarterly Rating     | Latest quarterly performance rating                      | Categorical | 1, 2, 3, 4  |