CS 457 - Homework Assignment 1: Data Type

rk06451

CS 457 Data Science
Habib University
Spring 2023

1. [30 points] List each attribute (column) given in the data set, choose its analytical data type (NOIR type) and explain why you consider this type. Answer without explanation will not be accepted.

Solution:

(A) **Time =** Interval, because time does not have "true-zero" as we cannot tell when does the time started thus, it is interval.

(B) **Latitude =** Ratio, latitude is continuous and quantitative and all mathematical operations are defined on that and can be 0 as well

(C) **Longitude =** Ratio, latitude is continuous and quantitative and all mathematical operations are defined on that and can be 0 as well

(D) **depth =** Ratio, it is measured in meter or any other S.I unit. It is continuous and quantitative and all mathematical operations are defined on that and can be 0 as well

(E) **mag =** Interval, it is quantitative quantity and mathematical operations can be performed on it. However, this lacks the "true-zero" property which means that magnitude does not exsist, thus this comes under Interval Scale.

(F) **magType =** Nominal, as ml and md are two qualitative properties which does not have any logical order. It can be two factors upon which the magnitude can be measured.

(G) **nst=** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(H) **gap =** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(I) **dmin =** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(J) **rms =** Ratio, Root Mean Square is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(K) **net =** Nominal, this column has values which does not have any logical order and are part of mutually exclusive.

(L) **id =** Nominal, this column has values which does not have any logical order and are part of mutually exclusive.

(M) **updated =** Interval, because this column have time for updating and time does not have "true-zero" as we cannot tell when does the time started thus, it is interval.

(N) **place =** Interval, this column has data which defines location with accurate distance reference to particular direction. This data can be used for addition, subtraction for pointing out the distance of one place to another.

(O) **type =** Nominal, this column has values which does not have any logical order and are part of mutually exclusive.

(P) **horizontal =** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(Q) **depthError =** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(R) **magError =** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(S) **magNst =** Ratio, it is quantitative quantity and operations such as Median, Mode, Multiplication can be performed on it.

(T) **status =** Nominal, as automatic and reviewed are two factors which only define whether anything has been human-reviewed or not. It does not have any logical order so this comes under Nominal scale.

(U) **Location =** Nominal, this column has values which does not have any logical order and define only the place.

(V) **magSource =** Nominal, This column has values which are part of set of mutually exclusive codes that does not contain any logical order and it can only have operations of equal and not equal, which are properties of Nominal.

2. 30 points Pick analytical data type (NOIR type) for each of the items below and explain why you consider this type. Answer without explanation will not be accepted.

   (a) Time with possible values AM or PM

Solution: Nominal
AM or PM does not have any logic order and are part of mutually exclusive codes and can have either equals or not equals to operation

(b) Brightness as measured by a light meter.

Solution: Ratio
Since Light meter is the part of measuring quantity which gives the result in light years. This is again quantitative and operations such as addition as multiplication can be performed on them.

(c) Brightness as measured by people's judgments.

Solution: Ordinal
This is based on the assumption that humans cannot accurately tell how bright the light is. They can just whether something is more brighter than other, and this eventually is an action of sorting. Thus equals and sorting can only be performed on them thatswhy they are part of Ordinal

(d) Angles as measured in degrees between 0 and 360

Solution: Ratio
Angles are countered as Continuous and quantitative and can be used for further mathematical operations such as finding the difference between two angles and etc

(e) Bronze, Silver and Gold medals as awarded at the Olympics.

Solution: Ordinal
The different medal tiers can be sorted according to their rank such as Bronze then Silver and then Gold so they come in category of Ordinal

(f) Height above sea level.

Solution: Ratio
Height is a continuous data and all the mathematical operations can be done on that. However, it can be Interval as well if we selected level is arbitrary

(g) Number of patients in a hospital

Solution: Ratio
Number of patients are quantitative and can have operations of multiplication, and all operations of interval. It also have "true-zero" property as well which means there can be an empty hospital as well, so it comes under category of Ratio

(h) ISBN numbers for books. (Look up the format on the Web)

Solution: Nominal
ISBN numbers are considered as string of numbers that have no logical order and makes the category an entire set of data

(i) Ability to pass light in terms of the following values: opaque, translucent, transparent.

Solution: Ordinal
opaque, translucent, and transparent are type of measurable parameters which decides how much light has to pass. In this way we can sort the data to check on which value the last passed least, thus this will come under category of Ordinal

(j) Military rank

Solution: Ordinal
The military rank can be sorted starting from the constable to chief or army staff which can further help to sort the entire Data.

(k) Distance from the center of HU campus.

Solution: Ratio
Distance from the center of HU campus is continuous and comes in interval and can be entitled for operations such as multiplication, addition and sort

(l) Density of a substance in grams per cubic centimeter.

Solution: Ratio
gram per cubic centimeter is knows as volume and volume is again continuous and quantitative and it has "true-zero" as well.

3. 30 points Assume you are doing a study of Habib University students' academic and demographic characteristics and storing this information as a dataset. Identify total of 12 attributes (3N, 3O, 3I, and 3R) and explain why you think is the correct data type. Answer without explanation will not be accepted

Solution:
Nominal:

1. **Student ID =** Student Id is the combination of strings of letters and numbers but cannot be sorted according to any meaningful scale. Thus only equal operations can be performed which makes it Nominal

2. **Student Gender =** Gender Identity such as Male(M) Female(F) or Other (O) does not have any logical Order and part of Mutual Exclusion
3. **Student Selected Degree =** Selected Degree means that a student can select Computer Science, Social Development Policy and Electrical Engineering Degree in which the order of value is not known.

Ordinal:

1. **Student Grade =** Student Grades can be A, B, C and D on which the sorting can be performed making A on first rank and D on the last rank.
2. **Designation =** This means on which designation student currently is, as student may be President of student council, any prefect or just normal student. This data can be sorted as well which will refine student categories.
3. **Class Ranking =** Class Ranking of student means on which scale student stands in class such as Excellent, Satisfactory, Average and Worst. In this way the student data can be sorted on these scales

Interval:

1. **GPA =** GPA of student is continuous type of quantitative data on which actions such as addition, subtraction can be performed. For example in one semester there may be more points and in one semester there may be few points.
2. **Age =** Age of Student is interval because in university usually everyone has age greater than 1 years. Here, age can be counted in intervals of 15-19, 20-25 and so on. Secondly, it does not have "true-zero" as well thus this comes under Interval Category
3. **Date of Birth =** Dates are Interval because it does not have "true zero"

Ratio:

1. **Tuition Fee =** Tuition fee of any student is quantitative data which can have "true zero" as well as student can have full scholarship as well.
2. **Income =** It can be possible student may be working as Teacher Assistant at Habib University and earning anything. Thus income is a quantitative quantity on which different operations can be performed.
3. **Number of Courses =** By this I mean, how many courses student has taken this semester, and since this is quantitative and can have "true-zero" value as may be student is freshie and has not enrolled in any of the course.