Zain Ul Haq

Zh05616

Data Science

Zeehasham Rasheed

HW 1

Part 1

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

Time:

(Sample value = 2023-09-04T10:25:56.930Z)

Ordinal. Time values cannot be added to one another so not interval or ratio. Values have a sequence, and a median can be found. Hence ordinal.

Latitude

(Sample value = -43.5513)

This is again an interval data type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we cannot multiply and divide, and time does not have a true zero either, as there are negative values included too so the possibility of there being a nonexistent zero does not exist.

Longitude

(Sample value = -116.7269974)

This is again an interval data type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a

median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we cannot multiply and divide, and time does not have a true zero either, as there are negative values included too so the possibility of there being a non existent zero does not exist.

Depth

(Sample value = 18.42)

This is again an interval data type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we cannot multiply and divide, and time does not have a true zero either, as there are negative values included too so the possibility of there being a nonexistent zero does not exist.

Mag

(Sample value = 0)

This is again an interval data type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we cannot multiply and divide, and time does not have a true zero either, as there are negative values included too so the possibility of there being a nonexistent zero does not exist.

MagType

(Sample value = ml)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

Nst

(Sample value = 20)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and time does not have a true zero either, as there are **NO** negative values included too and zero depth means that the nst does not exist, and as the data suggests, there are no negative values and hence, a zero would mean that the nst is NULL.

Gap

(Sample value = 106)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and time does not have a true zero either, as there are **NO** negative values included too and zero gap means that the gap does not exist, and as the data suggests, there are no negative values and hence, a zero would mean that the gap is NULL.

Dmin

(Sample value = 0.01)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and time does not have a true zero either, as there are **NO** negative values included too and zero dmin means that the dmin does not exist, and as the data suggests, there are no negative values and hence, a zero would mean that the dmin is NULL.

Rms

(Sample value = 0.13)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and as the data suggests, there are no negative values and hence, a zero would mean that the rms is NULL.

Net

(Sample value = nn)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

Id

(Sample value = nn00864170)

Ordinal. Values cannot be added to one another so not interval or ratio. Values have a sequence, and a median can be found. Hence ordinal.

Updated

(Sample value = 2023-08-08T22:16:40.030Z)

Ordinal. Values cannot be added to one another so not interval or ratio. Values have a sequence, and a median can be found. Hence ordinal.

Place

(Sample value = 46 km ESE of Beatty, Nevada)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

Type

(Sample value = earthquake)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

HorizontalError

(Sample value = 10.5)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and as the name suggests, 0 means that there is no error, and as the data suggests, there are no negative values and hence, a zero would mean that the horizontal error is NULL.

DepthError

(Sample value = 0.81)

As the name suggests, 0 means that there is no error. Hence,

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and as the name suggests, 0 means that there is no error, and as the data suggests, there are no negative values and hence, a zero would mean that the depth error is NULL.

MagError

(Sample value = 0.16)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and as the name suggests, 0 means that there

is no error, and as the data suggests, there are no negative values and hence, a zero would mean that the mag error is NULL.

MagNst

(Sample value = 4)

This is now a ratio type, as along with having frequency of distribution and the possibility of finding a mode (nominal), and a specified order with the possibility of finding a median (ordinal), we can add or subtract the values and quantify the difference between each value (interval). Moreover, we **CAN** multiply and divide, and as the data suggests, there are no negative values and hence, a zero would mean that the nst is NULL, and as the data suggests, there are no negative values and hence, a zero would mean that the magnst is NULL.

Status

(Sample value = reviewed)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

LocationSource

(Sample value = hv)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

MagSource

(Sample value = hv)

Nominal. There is no order of values. We cannot even find a median, so it is not ordinal. If it is not ordinal, it is none of the other data types. Counts can be done, so we can find the mode and Frequency of Distribution, hence only nominal.

Part 2

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.

1. Time with possible values AM or PM.

Ordinal. Can be sequenced so they are ordinal. 1am will come ahead of 12am. Median of times from a list of times can be found. However, since they cannot be added to one another, they are not interval data type.

2. Brightness as measured by a light meter.

Quantifiable values that can be added or subtracted along with a sequence and zero brightness would mean no brightness. Hence, ratio type.

3. Brightness as measured by people's judgments.

Non quantifiable values but still a comparison can be made between two values and a sequence can be made. Hence, ordinal.

4. Angles as measured in degrees between 0º and 360º.

Ratio data type. 0 degrees actually means a nonexistent angle. Other than this, the intervals are well defined, and the data is comparable. You can add and subtract degrees too.

5. Bronze, Silver and Gold medals as awarded at the Olympics.

Ordinal. They can be compared to one another but are not quantifiable and don't have fixed intervals. Values cannot be added or subtracted.

6. Height above sea level.

Ratio data type. 0 height would be nonexistent and hence, have some special significance. Other than this, values would be easily sequenced and added/subtracted to one another.

7. Number of patients in a hospital.

The number of patients in different hospitals can be compared with other hospitals. 0 patients would mean the nonexistence of patients. Numbers can be added or subtracted to each other as well as multiplied or divided. Hence, ratio.

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

Can be compared to one another. All ISBN values are different from each other. Data can be compared and ordered. However, values cannot be added or subtracted and we cannot quantify the difference between values. Hence, ordinal.

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

Not comparable to one another. Can be counted. A frequency of distribution can be made. Hence, nominal.

10. Military rank.

Ordinal. Values are not quantifiable, hence not interval or ratio but can be compared.

11. Distance from the center of HU campus.

Value cannot be negative and hence a zero would signify the nonexistence of distance from the center. Values can be added/subtracted and multiplied/divided. Easily comparable. Easily quantifiable. Hence, ratio.

12. Density of a substance in grams per cubic centimeter.

Value cannot be negative and hence a zero would signify the nonexistence of distance from the center. Values can be added/subtracted and multiplied/divided. Easily comparable. Easily quantifiable. Hence, ratio.

Part 3

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.

For example: *Academic Year: (Freshman, Sophomore, Junior, Senior) = Ordinal* (because they can be ordered based on their admission year)

3N:

Financial Hold: (False, True) = nominal (because this information cannot be ordered, but a count can be made of all False/True values)

Academic standing: (Good Academic Standing, Bad Academic Standing) = nominal (because this information cannot be ordered, but counts can be made and a frequency of distribution can hence be made)

Disciplinary Conduct: (True, False) = nominal (because this information cannot be ordered, but a count can be made of all False/True values)

3O:

ID: (aa00000-zz99999) = ordinal (because this information can be ordered alphabetically then numerically, but not added/subtracted or quantified)

Email: (id@st.habib.edu.pk) = ordinal (because this information can be ordered with respect to ID)

ECA performance: (Very good, good, satisfactory, poor, very poor) = Values can be ordered into 1,2,3,4,5 but not added/subtracted to one another. Hence, ordinal

3I:

Graduating year: (yyyy) = interval (because year can be 0000 too possibly and that doesn't negate the existence of a year. However, values can be multiplied/divided and quantified. A sequence can be made.

Year of Arrival (yyyy) = interval (because year can be 0000 too possibly and that doesn't negate the existence of a year. However, values can be multiplied/divided and quantified. A sequence can be made.

Cafeteria credit (-infinity - +infinity) = interval (if we assume we can store credit in cards and take loan from cafeteria in case of not having money at the given time. This value does not become nonexistent if it is zero. Zero is also a valid value. Both negative and positive values can exist, and values can be added and subtracted, even sequenced/ordered.)

3R:

Number of books issued from library: (0-infinity) = interval (because this information can be quantified, and sequenced as well as multiplied/divided)

GPA (0.0-4.0) = ratio because 0 means nonexistence of GPA, but it can be quantified, and values can be multiplied/divided, and sequenced.

Number of courses taken this semester (0-7) = ratio because 0 means nonexistence of courses, but it can be quantified, and values can be multiplied/divided, and sequenced.