**Data Science – Homework 1: Data Types**

Ali Hashir – ah05433

**Part 1**

List each attribute (column) given in the dataset, choose its analytical data type (NOIR type) and explain why you consider this type.

|  |  |  |
| --- | --- | --- |
| **Attribute** | **Data Type** | **Reasons** |
| Time | Intervals | It does not have a true zero value, such as the time is 00:00:00 indicates that it is midnight rather than implying that time does not exist. |
| Latitude | Intervals | It does not have a true zero value, 0 latitude does not mean that latitude does not exist. |
| Longitude | Intervals | It does not have a true zero value, 0 longitude does not mean that longitude does not exist. |
| Depth | Intervals | It does not have a true zero value. |
| Mag | Ordinal | There is a natural order of magnitude, we don’t know the exact difference but we know 4.0 is more severe than 3.0. |
| Magtype | Nominal | These are just categorical variables with no particular order. |
| NST | Ratio | It is a quantity with a true zero value. An earthquake can occur without anyone feeling it, so the NST attribute can be zero. |
| Gap | Ratio | Contains true zero value, when there is no gap in between. |
| Dmin | Ratio | Contains true zero value. |
| RMS | Ratio | Contains true zero value. |
| Net | Nominal | These are just categorical variables with no particular order. |
| Id | Nominal | It is a unique identifier for each earthquake. It is a string of characters that does not have any inherent meaning. |
| Updated | Intervals | It does not have a true zero value. |
| Place | Nominal | These are just categorical variables with no particular order. It is a label for the epicenter of the earthquake. |
| Type | Nominal | These are just categorical variables with no particular order. |
| Location source | Nominal | These are just categorical variables with no particular order. |
| Mag source | Nominal | These are just categorical variables with no particular order. |
| Horizontal error | Ratio | Contains true zero value, when there is no error. |
| Depth error | Ratio | Contains true zero value, when there is no error. |
| Mag error | Ratio | Contains true zero value, when there is no error. |
| Mag NST | Ratio | Contains true zero value, if the earthquake has not been felt by anyone. |
| Status | Nominal | These are just categorical variables with no particular order. |

**Part 2**

Pick analytical data type (NOIR type) for each of the items below and explain why you consider this type.

|  |  |  |
| --- | --- | --- |
| **Data** | **Data type** | **Reason** |
| Time with possible values AM or PM. | Nominal | The time attribute has two categories, AM and PM, that cannot be ordered logically. This makes it a binary nominal type. |
| Brightness as measured by a light meter. | Ratio | The light meter has a clear definition of zero. Zero brightness means that there is no light. |
| Brightness as measured by people’s judgments. | Ordinal | People will rate the light intensity by using variables such as least, less, more, or most. These variables can be used to compare and order the brightness of different light sources. |
| Angles as measured in degrees between 0ᵒ and 360ᵒ. | Ratio | They have a meaningful zero point, a continuous range, and no negative values. |
| Bronze, Silver and Gold medals as awarded at the Olympics. | Ordinal | Medal is a discrete quantity, that can be ordered and compared among themselves. It can be ranked as first, second, or third based on the medal. |
| Height above sea level. | Ratio | It is a numeric value that can be measured and compared to other values. It has a zero point at sea level. |
| Number of patients in a hospital. | Ratio | Hospital patient numbers can be compared and there can be zero patients in a hospital. |
| ISBN numbers for books. | Nominal | It does not have any logical order, instead it is used to label or for unique identification of the book. |
| Ability to pass light in terms of the following values: opaque, translucent, transparent. | Ordinal | It can be ordered among themselves. It can be compared how an object is less or more transparent. |
| Military rank. | Ordinal | There is a hierarchy of authority and responsibility that is based on rank. As one rises through the ranks, they are given more authority and responsibility. |
| Distance from the center of HU campus. | Ratio | Distance from the HU campus center is a continuous numerical number with a  relevant zero point. The distance is zero when we are in the center of HU campus |
| Density of a substance in grams per cubic centimeter. | Ratio | Density can also be absolute zero. |

**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.

**Nominal:**

* Name: It does not have any logical order. It contains the names of the student.
* Gender: It has exactly three or more mutually exclusive categories F, M or Other, which does not contain any logical order.
* City: Student can belong to any city and there is no logical order to them.

**Ordinal:**

* Age: It can be ordered among themselves. The values can be compared whether one person is younger than the other or not.
* Batch: It can be ordered or ranked among themselves. The values can be compared whether a student is a sophomore or a senior.
* Academic Standing: It can be ordered or ranked among themselves. Whether a pupil is doing well academically or not, the values can be compared.

**Interval:**

* Date of Birth: They are continuous measurements of a roughly linear scale.
* GPA (Grade Point Average): It is a quantitative measure of academic performance that can take on any value within a certain range. The GPA scale is continuous and has equal intervals, which means that the difference between any two GPAs is the same.
* Enrollment date: They are continuous measurements of a roughly linear scale

**Ratio:**

* Financial aid in Percentage: It has a clear definition of ‘zero’. When there is zero percentage then it means that the student has not received any financial aid.
* Attendance: Attendance is a numerical value that has a meaningful zero point. A zero attendance value means that the person was not present.
* Friends in Habib: The number of Friends a student has in Habib, is a discrete numerical attribute with a true zero point. It can be 0 if the student has no friends.