**COIS 4400H Assignement 1 Q1**

**Q1 . Categorical to numerical representations**

For each part I have explained my choice of mapping used with the reason why I chose to do that.

**a. Colors of rainbow**

Colors of a rainbow is ordinal data. In order for the colors to make a rainbow, they have to be in an order otherwise they are just a group of colors. Therefore, using the m of n mapping
This way, the order is maintained, and values are allocated as evenly as possible while maintain that given order.

| | **Old value** | **New values** | | |
|---|---|---|---|---|
| | **Color** | **Attribute 1** | **Attribute 2** | **Attribute 3** |
| **Sample 1** | Violet | 0 | 0 | 0 |
| **Sample 2** | Indigo | 0 | 0 | 1 |
| **Sample 3** | Blue | 0 | 1 | 0 |
| **Sample 4** | Green | 1 | 0 | 0 |
| **Sample 5** | Yellow | 0 | 1 | 1 |
| **Sample 6** | Orange | 1 | 0 | 1 |
| **Sample 7** | Red | 1 | 1 | 0 |

**b. Names of employees**

This data can be considered in 2 ways. One – Nominal – whether or not the person is an employee and hence whether or not it can be considered as an employee name. The second way is – ordinal – in that case, all the names of the employees must be stored in an alphabetical way (like in any ledger). In that case we consider the alphabetical start of their name and then carry on that way.
I am considering it a simple nominal data – a yes, is an employee and a no, is not an employee.

| | **Old value** | **New values** | |
|---|---|---|---|
| | **Is Employee** | **Employee Yes** | **Employee No** |
| **Sample 1** | Yes | 1 | 0 |
| **Sample 2** | Yes | 1 | 0 |
| **Sample 3** | No | 0 | 1 |
| **Sample 4** | Yes | 1 | 0 |
| **Sample 5** | No | 0 | 1 |

**c. Which continent a country belongs to**

This would be a Nominal since there is no order when it comes to continents (thankfully). So for knowing which continent a country belongs in we just need a 1 of n mapping. So since there are 7 continents (yes, Antarctica as well because someone actually said it is not considered a continent? I don't know, google says yes and I was taught yes in school), we need 7 attribute values as new columns for these attributes. Now I know Antarctica has no countries and it can not be considered as an attribute, but since it is still a continent I would consider it as one of the attributes.

| | **Old value** | **New values** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Continent** | **Asia** | **Europe** | **Oceania** | **N America** | **S America** | **S Africa** | **Antarctica** |
| **Canada** | North America | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| **India** | Asia | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Morocco** | South Africa | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| **Italy** | Europe | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| **New Zealand** | Oceania | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| **Brazil** | South America | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

**d. Letters of the alphabet**

Letters of alphabets – ordinal : there is a certain hierarchy to it, start with 'a' and ends with a 'z'. Now, using the tabular mapping methods as m of n mapping would not be very efficient since there would be 26 attributes. A question to ask can letter of alphabets be considered a repetitive data like months of a year? I think, yes, because no matter how many years pass, the letters of alphabets will always be 26, which means we start back from 'a' when we end at 'z'. Therefore, a cyclic representation would be a good way.