A first introduction to important data terms Suggested solution

**Task 1**

Add five more *characteristics* and their *characteristic values* to the table.

|  |  |
| --- | --- |
| **Features in the YOU-PB data** | **Characteristic values** |
| Gender | female, male |
| Age | 10, 11, 12, 13, ..., 20 |
| Grade level | 5, 6, 7, ..., 13 |
| Ebook\_read | Never, less often, once a month, once a fortnight, once a week, several times a week |
| Type of school | Gymansium, Realschule, Hauptschule, Gesamtschule, other |
| Fixed games console available | Available at home, not available at home |
| Play console | Never, less often, ..., several times a week, daily |

**Task 2: Categorical or numerical characteristics**

1. Classify the following characteristics from the YOU-PB dataset into the categories "categorical" or "numerical". Briefly explain your choice.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Categorical or numerical?** | **Reason** |
| Gender | Categorical | There are two different characteristics (categories) |
| Age | Numerical | Can be sorted in ascending order / only numbers occur as characteristics |
| Accounts\_number | Numerical | Can be sorted in ascending order / only numbers occur as characteristics |
| Podcasts\_listen | Categorical | There are different characteristics (categories) |
| Instagram\_use | Categorical | There are different characteristics (categories) |
| Gaming\_Minutes | Numerical | Can be sorted in ascending order / only numbers occur as characteristics |
| eReader\_present | Categorical | There are different characteristics (categories) |
| Public\_transport\_APP\_use | Categorical | There are different characteristics (categories) |

1. Why are there probably many more categorical than numerical characteristics in this survey?

Because this survey mainly asked how often something is used and the corresponding answers are distributed categorically.