Coding Assignment

A school has conducted a survey on the students’ feelings on the use of their mother tongue in daily life.

Twelfth questions were included in the survey. For a prelim study, you are tasked to analyse the first eight questions.

The eight questions are

* When I play games in <my mother tongue> I feel?
* When I hear a song in <my mother tongue> I feel?
* When my teacher speaks to me in <my mother tongue> I feel?
* When my parent read to me in <my mother tongue> I feel?
* When I do worksheet in <my mother tongue> I feel?
* When I talk to my friend in <my mother tongue> I feel?
* When I watch TV in <my mother tongue> I feel?
* When I learn a new word in <my mother tongue> I feel?

<my mother tongue> can be substituted with Mandarin, Malay or Tamil.

The rating for each question are:

|  |  |
| --- | --- |
| Rating |  |
| 2 | Very Happy |
| 1 | Happy |
| 0 | Neutral |
| -1 | Dislike |
| -2 | Extreme dislike |

You are to perform the following tasks:

1. Data cleaning (50%)

Validate all the entries.

* Ensure no blank rating entry (blank entry)
* Ensure no invalid entries, for example letters (invalid rating entry)
* Ensure all rating entries are in the correct range (invalid rating range)

Remove all invalid entries.

Remove entry with Subject ID “test”.

Remove all the other questions response entry.

Present the script that you have used to validated and remove the entries

Name your script as data\_cleaning.py

State the entries (subject ID and language) that you have removed with the reason for

removal in a report. (e.g. P123,Mandrain, blank entry)

The possible reasons are: blank entry, invalid rating entry and invalid rating range.

No need to state the Subject ID “test” removal.

Store the cleaned data in a new csv file and name it as cleaned\_data.csv.

1. Data Analysis (50%)

Using the cleaned data csv file:

* 1. For each question, compute the overall
* Mean
* Sample Standard Deviation
  1. For each mother tongue and question, compute
* Mean
* Sample Standard Deviation
  1. For each mother tongue, compute the 8 questions overall
* Mean
* Sample Standard Deviation

Present the three scripts to generate the above computation.

Name your script as partb\_1.py, partb\_2.py and partb\_3.py