



Tidy Data

Classroom Activity 4

Objective

The student will perform data preprocessing for further analysis.

Instructions

1. Identify the variables in the following datasets based on the Tidy Data Philosophy:
 - a. Relationship between income and religion in the US from the Pew Forum.
 - b. Billboard Dataset.
 - c. Tuberculosis Dataset from the World Health Organization.
 - d. Weather data from the Global Historical Climatology Network in Mexico.
 - e. Rank of Baby Names from Illinois in 2014 and 2015.
2. Clean the datasets and convert them into Tidy Data.
3. Identify the type of error in each dataset. Write it in the notebook.

Homework Submission

Submit the following:

- Tidy Data – Notebook (ipynb).

Teams

Classroom Activity 4 can be done in teams.

Deadline

- October 16th, 2019 at 12:00 im.

Suggested Readings:

- Tidy Data (Hadley Wickham)
<https://vita.had.co.nz/papers/tidy-data.pdf>



Grading Scheme

Results	Excellent	Good	Fair	Needs Improvement
Notebook				
The students are able to apply correctly the Tidy Data Principles: identifying the type of error and cleaning the Dataset 1.	20.0 pts	15.0 pts	10.0 pts	5.0 pts
The students are able to apply correctly the Tidy Data Principles: identifying the type of error and cleaning the Dataset 2.	20.0 pts	15.0 pts	10.0 pts	5.0 pts
The students are able to apply correctly the Tidy Data Principles: identifying the type of error and cleaning the Dataset 3.	20.0 pts	15.0 pts	10.0 pts	5.0 pts
The students are able to apply correctly the Tidy Data Principles: identifying the type of error and cleaning the Dataset 4.	20.0 pts	15.0 pts	10.0 pts	5.0 pts
The students are able to apply correctly the Tidy Data Principles: identifying the type of error and cleaning the Dataset 5.	20.0 pts	15.0 pts	10.0 pts	5.0 pts