## Module 3: Data Acquisition, Evaluation and Exploration

Assignment

# edureka!

edureka!

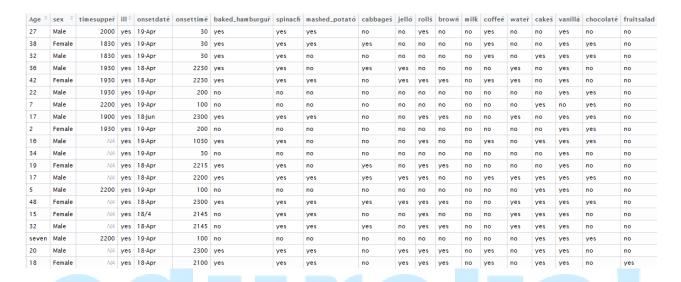
© Brain4ce Education Solutions Pvt. Ltd.

### Module 3: Assignment

Given a dataset from an outbreak of food-borne illness in the US in 1940.

The dataset can be downloaded from the link:

https://edureka.wistia.com/medias/qp5fzq03hd



#### Column information:

- age the person's age in years
- sex the person's gender
- timesupper the time the person ate (nearest half hour)
- ill whether the person developed GI illness after the supper
- onsetdate the date of onset of illness for those who became ill
- onsettime the time that the person reported first feeling ill (nearest half hour)
- 15 variables indicating whether the person reported eating specific food items at the supper

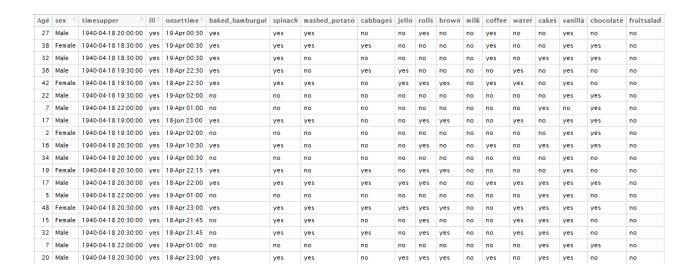
#### Perform the following tasks on the dataset

#### Task 1: Import the dataset into R

→ Ensure imported data is interpreted correctly by R.

#### Task 2: Perform Data Wrangling and clean the data

- → Convert Age column to type Numeric
- → Omit NULL values from non-numerical data fields
- → Replace NULL values with average value of that variable
- → Give proper time format for timesupper column using strptime() and get the date and time in the below specified format
- → Final output should be like



#### Task 3: Perform Analysis and Visualization using appropriate type of graph, to find out:

- → Which is the most consumed food item among the patients.
- → Find the average age of people who are ill using Boxplot.
- → Visualize gender ratio from the data.