**Principles of Data Science (5530)-Assignment 1**

**Name: Praveen Kumar Reddy Kadapala**

**ID: 16355737**

**1. Answer:**

Stage 1 – Data collection:

Raw Data is gathered or fetched and stored in this stage.

**Folder structure:**

A black screen with a blue line

Description automatically generated

* Acquired data is stored in the form of .csv file so that the analysis can be done.
* README.txt file contains the metadata about the dataset.

A white background with black text

Description automatically generated

Stage 2 – Data Processing:

Here, gathered or fetched data will be preprocessed so that it will be ready and good for analysis.

A screenshot of a computer

Description automatically generated

* Several preprocessing operations are carried out, like renaming the columns with appropriate names and converting the categorical values to numerical values and then the final obtained clean data is kept in a separate file for performing analysis.

**Folder Structure:**

**A computer screen with white text

Description automatically generated**

**Stage 3 – Data Analysis:**

* To find the relationships or patterns between the fields or observations of the dataset, data analysis is performed on the dataset by running the scripts.

**A screenshot of a computer screen

Description automatically generatedA screen shot of a computer screen

Description automatically generated**

* At this level, analysis is done on the cleaned data.
* Below are the plots of two different analyses performed on the frailty dataset.
* A box plot graph is created with data of grip strength on the Y-axis and frailty on the X-axis.

A diagram of a relationship between grip strength and frally

Description automatically generated

**Analysis Result:** The box plot clearly shows that the persons who has frailty symptoms has less grip strength compared to the persons not having frailty symptoms.

* A box plot graph is created with data of Age on the Y-axis and frailty on the X-axis.

A diagram of a relationship between age and fracturing

Description automatically generated

**Analysis Result:** The box plot clearly shows that the people are observing or having frailty symptoms only in the age between 22 to 35. This indicates that frailty symptoms are observed mostly in the age between 22 – 35.

**Folder Structure:**

**A screenshot of a computer

Description automatically generated**