4/7/2024

ECPI University

Angeline Ortiz de los Santos

1.8 Project

Data Set Selection

* How many patients are included in the dataset?
  + 15,001 are included in the data set.
* What percentage of patients have been diagnosed with a stroke?
  + 7,469 patients have been diagnosed with a stroke.
* How many unique work types are there in the dataset?
  + There are 4 unique work types, and these are government job, never worked, private, and self-employed.
* What is the age distribution of the patients?

Patients age falls between 18 years old to 90 years old.

* Are there any missing values? If so, in which columns?
  + No blank cells were found.
* Identify at least three potential questions or hypotheses that could be addressed using this dataset.
  + Potential of patient having a stroke.
  + Trend of Lifestyle such as dietary, alcohol smoking can provoke a stroke based on data collected?
  + Does family history play a role in strokes based on data collected?