

Jeremy Chen

The 2014 Inc. 5000

This dataset is a secondary source that is collected to be observed for researchers. The data is formatted in a .csv file that has structured tables with different data types and variables. This data type contains integers, strings, and characters. Most of the variables are independent variables that do not depend on other factors. People who would be interested in this data set are most likely trying to invest into businesses that have great growth potential. In my opinion this data set is very detailed on general information about a company. It is like an overview of all 5000 companies so people can quickly scan through them for their benefit.

Stock Exchange Data

This dataset is a secondary source collected from Yahoo finance to track stock exchanges from all over the world. (US, China, Canada, Germany, Japan, etc) It was updated three months ago and the author is a Digital Marketing Consultant at Burst Analytics. This is also a .csv file with a table that has different data types and variables in each cell that is structured. The data types presented are dates, integers, floating points, and strings. These are independent variables that have their own values for each record. Those who would use this are people who are planning to invest in stocks. Therefore they need to do a lot of research beforehand. I think this data set can be more detailed. It looks like it is just a short brief overview of the stocks.

Business and industry report

This dataset is a primary survey report from a US census. This data is the version that was updated 4 years ago. The US Census Bureau created this data to cover 16 of their economic data reports and surveys. It is in a table with structured columns that have values in a .csv file. The data types that are given in the file are dates, strings, and integers. The value variable is dependent in order to show the reasonable units we need to drop entries that are equal to the string less than 0.05 percent. The rest of the variables are all independent that are manually entered. This is a census report that stretches back into the 19th century. People can use this to see the growth of business reports over time. In my opinion this is a large data set that is collected over time. It is a very vague data set with information that is confusing to people. It only represents the time series, date and value.