How to know if you have the right data

Data Analysis
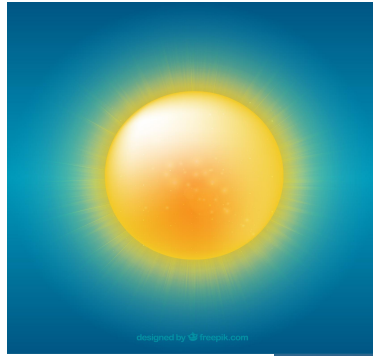
Number of observations is too small
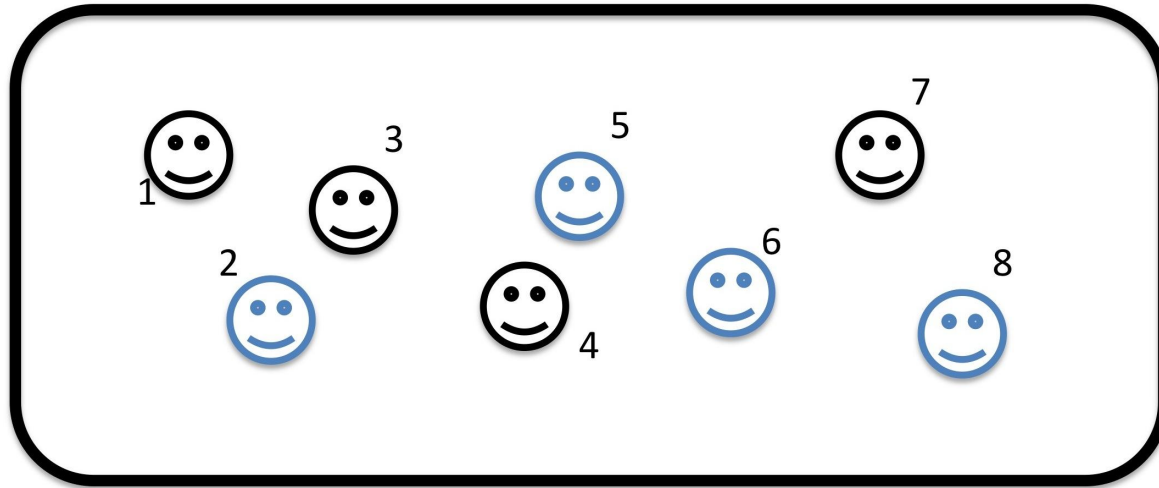
Dataset does not contain the exact variables you are looking for

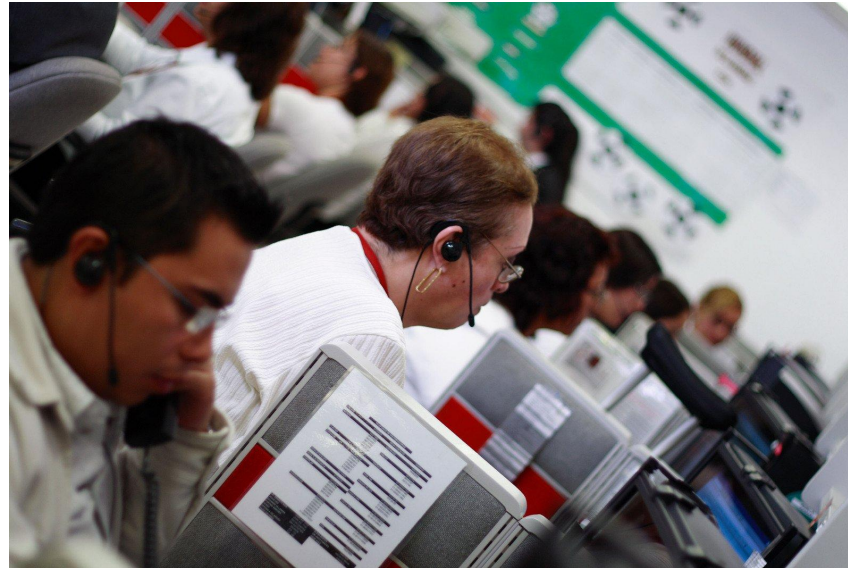# Variables in the dataset are not collected in the same year

Dataset is not representative of the population that you are interested in

Dataset is not representative of the population that you are interested in

Dataset is not representative of the population that you are interested in

Examples of bad sampling:

- Surveying subscribers of a gun-related magazine for research on attitudes toward owning guns
- Surveying Facebook users for what TV shows Americans like

Best sampling practices:

- Always think about what your population is
- Collect data from a sample that is representative of your population
- If you have no choice but to work with a dataset that is not collected randomly and is biased, be careful not to generalize your results to the entire population

Some variables in the dataset is measured with error

FiveThirtyEight

Politics   Sports   Science & Health   Economics   Culture

NOV. 9, 2016, AT 4:53 PM

The Polls Missed Trump. We Asked Pollsters Why.

By Carl Bialik and Harry Enten
Filed under 2016 Election

POLITICS

4 Possible Reasons The Polls Got It So Wrong This Year

November 14, 2016 · 2:28 PM ET

TheUpshot

POLITICAL CALCULUS

A 2016 Review: Why Key State Polls Were Wrong About Trump

By Nate Cohn

May 31, 2017

193

BUSINESS INSIDER

TECH   FINANCE   POLITICS   STRATEGY   LIFE   INTELLIGENCE   ALL

A group of major pollsters just released an autopsy report to explain why the polls were such a disaster in 2016
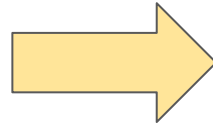
Allan Smith May 7, 2017, 12:23 PM

Image: freepik.com and Wikimedia Commons (Kyle Taylor)

Garbage in, garbage out!

Data Analysis