

Drill

In each of the scenarios, find possible shortcomings of the data sources to answer the given question.

What could be done to either adjust the analysis or reframe the question so that you can answer it accurately?

- Data Source: Amsterdam availability data scraped from AirBnB on December 24th.

Question: What are the popular neighborhoods in Amsterdam?

- The question is poorly defined. What's the meaning of popular? Popular to live in? Popular for tourists? Popular for brunch? I would suggest reframing the question so that you are asking what are the most popular neighborhoods for AirBnB tourism in Amsterdam.

- Data Source: Mental health services use on September 12, 2001 in San Francisco, CA and New York City, NY. Question: How do patterns of mental health service use vary between cities?

- Data from the day after 9/11 is likely to be skewed in NYC due to exceptional circumstance so one would not be able to confidently extrapolate general patterns. I would choose a data set from a different day more likely to be representative.

- Data Source: [Armenian Pub Survey](#). Question: What are the most common reasons

Armenians visit local pubs?

- This may be an exceptional circumstance. Are there only Armenians in these bars? Are we extrapolating to diaspora Armenians as well? I would rephrase the question to be more limited in scope.