Drill

In each of the scenarios, find possible shortcomings of the theoretical or actual 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?

**Theoretical Data Source: Amsterdam availability data scraped from Airbnb on December 24th.**

**Question: What are the popular neighborhoods in Amsterdam?**

The data is biased and unrepresentative of Amsterdam. First, Airbnb is not representative, excluding hotel. Popular neighborhoods may not correlate with Airbnb very well. Second, data was collected on a special event—Christmas eve. People may do things differently on that day. To adjust data quality and bias, I would collect one-year data from different sources, for example, housing websites.

**Theoretical Data Source: Mental health services use on September 12,**

**2001 in San Francisco, CA and New York City, NY.**

**Question: How do paterns of mental health service use vary between cities?**

Time of data collection is problematic. The day right after 911, there could be a peak of mental health service use in NYC. To adjust bias, I would collect longitudinal data in one year or over multiple years.

**Actual Data Source: Armenian Pub Survey.  
Question: What are the most common reasons Armenians visit local pubs?**

Armenians may not only visit Armenian pubs. To adjust data quality, I would collect data from local pubs among Armenians.