Exercises in Probability

* Calculate the probability of flipping a balanced coin four times and getting each pattern: HTTH, HHHH and TTHH. 3/13

**HHHH**

HHHT

HHTH

HTHH

THHH

HHTT

**HTTH**

**TTHH**

TTTH

TTHT

THTT

HTTT

TTTT

* If a list of people has 24 women and 21 men, then the probability of choosing a man from the list is 21/45. What is the probability of not choosing a man?

24/45

* The probability that Bernice will travel by plane sometime in the next year is 10%. The probability of a plane crash at any time is .005%. What is the probability that Bernice will be in a plane crash sometime in the next year?

.0005%

* A data scientist wants to study the behavior of users on the company website. Each time a user clicks on a link on the website, there is a 5% chance that the user will be asked to complete a short survey about their behavior on the website. The data scientist uses the survey data to conclude that, on average, users spend 15 minutes surfing the company website before moving on to other things. What is wrong with this conclusion?

The data scientist should not use this data to come to conclusions about the population of users of the company website. The longer the user stays on the website the more likely they will be asked to complete the survey. Therefore it is an inaccurate count of the length of time spent on the website. The survey skews towards having responses from users who have spent a greater amount of time on the website.

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?

* **Data Source**: Amsterdam availability data scraped from AirBnB on December 24th. **Question**: What are the popular neighborhoods in Amsterdam?

Shortcomings - The data was scraped on Christmas eve therefore it is not representative of a typical day or time. The data is from a single sample. Clients of Airbnb might find particular neighborhoods more appealing than people who are not clients of Airbnb. The question could be reframed to ask According to Airbnb users what are the popular neighborhoods in Amsterdam during Christmas.

* 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?

Shortcomings – Because of 9/11 a very unusual event, mental health services most likely were not distributed in typical manner and were over burdened and shifted due to the tragic events in NYC. Therefore it is not a good comparative time frame or location. The analysis could be adjusted by picking another day or time frame for data analysis. Or the question could be modified to ask specifically about the effects of 9/11.

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

The most common reasons Armenians visit local pubs are to hang out with friends, the environment, the music and the food served.