

Name _____ Period _____

Skill 0.1 Exercise 1

What is data literacy?

Data literacy is the ability to understand, analyze, and communicate with data effectively

Why is data literacy important?

Data literacy helps us make decisions. Being data literate allows individuals and organizations to:

Make better decisions

Identify opportunities

And adapt to a data-driven world

Skill 0.2 Exercise 1

Provide another example of Garbage in, garbage out.

Any situation that omits data...

Election – not including everyone

Facial recognition software

Skill 0.3 Exercise 1

Neural Machine Translation (NMT) is trained on example text that exists in the world. Consider the Google Translate that was constructed using NMT.

The screenshot shows a Google Translate interface comparing Turkish and English. The source language is set to "Turkish - detected" and the target language is "English". The input text on the left is: "o bir aşçı", "o bir mühendis", "o bir doktor", and "o bir hemşire". The output text on the right is: "she is a cook", "he is an engineer", "he is a doctor", and "she is a nurse". There are small icons for microphone and speaker next to each word pair.

o bir aşçı	she is a cook
o bir mühendis	he is an engineer
o bir doktor	he is a doctor
o bir hemşire	she is a nurse

What is the bias in this translation.

Professions are sexist.

How might Google Engineers modify the translation to remove bias?

Provide a more diversified training set for each profession

Name _____ Period _____

Skill 0.4 Exercise 1

In 2014, Amazon experimented with using software to screen job applicants. The screening software was trained on a decade of résumés that had been previously rated by employees as part of the hiring process. Below shows the distribution of male and female employees at Amazon in 2014.

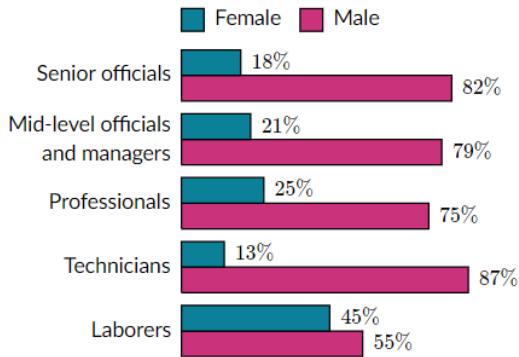


Chart source: [Seattle Times](#)

Is it statistically possible that the above results were by chance?

No

How might Amazon better train their software model to avoid “reckless regard” for the rights of female candidates?

Adjust the algorithm, or train with a more diversified set of data

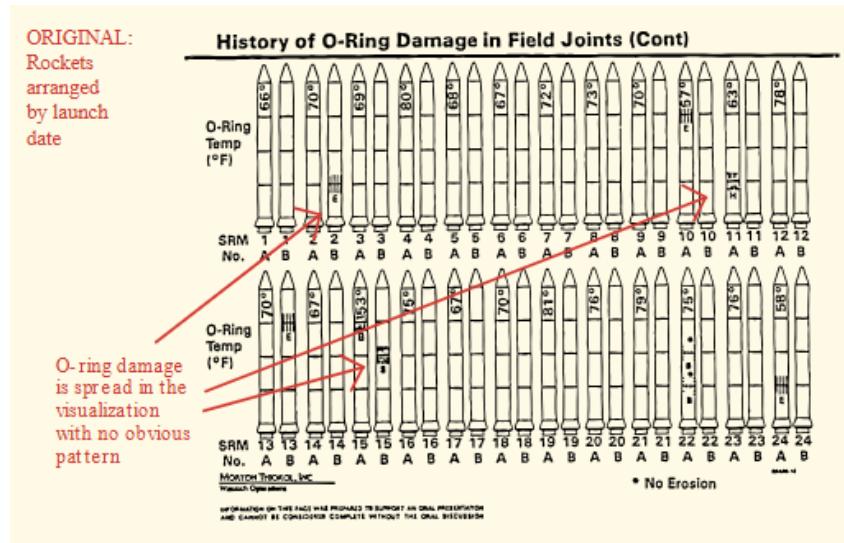
Skill 0.5 Exercise 1

Data Science with Python
Ticket Out the Door
Set 0: Visualizing Data Key

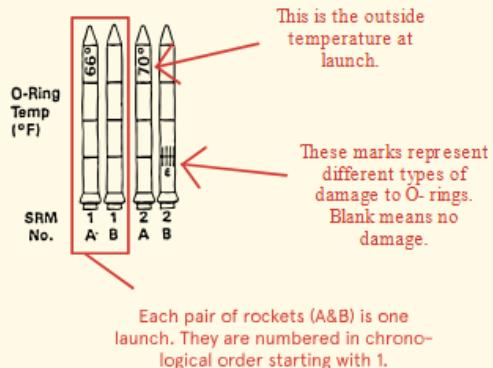
Name _____ Period _____

On January 28, 1986 the Challenger space shuttle carried seven US astronauts who were supposed to deploy a satellite and study Halley's Comet while they were in orbit. Less than two minutes after takeoff, however, the shuttle exploded, killing all seven crew members. The explosion was caused by a failure of two O-rings: small rubber rings that helped create an airtight seal between the space shuttle and its launch fuel supply. Before the launch, engineers were concerned about how the low-temperature forecast would affect the O-rings' ability to make a proper seal. The engineers made their arguments in favor of postponing the launch using, in part, a series of data visualizations that showed launch success rates at various temperatures. Tragically, their arguments did not prevent the launch from proceeding.

Below are the visualizations.



How to read the charts:



How could the engineers have improved their visualization, and potentially prevented this tragedy?

Data Science with Python Ticket Out the Door Set 0: Visualizing Data Key

Name _____ Period _____

