DATA SCIENCE DANGER ZONE

(1/1 point)

Why is a hacker with substantive expertise a dangerous combination? Select all answers that match:

Top of Form

Can end up performing statistically incorrect analyses, May fail to ignore irrelevant factors (more computation), Can create what legitimate appearing analysis without any understanding of how they got there or what they have created, - correct

Can apply Machine Learning to big data problemsCan end up performing statistically incorrect analyses May fail to ignore irrelevant factors (more computation) Uses traditional research to solve problemsCan create what legitimate appearing analysis without any understanding of how they got there or what they have created

Bottom of Form

Checkyour answer **Hide Answer**

(Optional reading) This interview study characterizes the process of industrial data analysis and documents how organizational features of an enterprise impact it: [Enterprise Data Analysis and Visualization: An Interview Study](http://idl.cs.washington.edu/files/2012-EnterpriseAnalysisInterviews-VAST.pdf). The study is based on semi-structured interviews with 35 data analysts from 25 organizations across a variety of sectors, including healthcare, retail, marketing and finance.

(Optional reading) This Harvard Business Review article, [Data Scientist: The Sexiest Job of the 21st Century](https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1), presents a deep dive on what organizations need to know about data scientists: where to look for them, how to attract and develop them, and how to spot a great one.

CURATING AND FILTERING DATA

(1/1 point)

Why is curating/filtering data a key model component?

Top of Form

Real-world data patterns are easy to modelReal-world data is very dirty Real-world data is very dirty - correctReal-world data should never be filteredReal-world data does not require curating

Bottom of Form

**EXPLANATION**

Real-world data often contains inconsistencies or errors that have to be filtered out

Checkyour answer **Hide Answer**

OVERCOMING ASSUMPTIONS

(1/1 point)

Why is overcoming assumptions hard? Select all the answers that match:

Top of Form

It is human nature to make assumptions, Humans have biases based on prior observations , - correct

Assumptions are always correctIt is human nature to make assumptions Humans have biases based on prior observations

Bottom of Form

**EXPLANATION**

We all naturally try to generalize when we see what appears to be a pattern. Also, once we observe a pattern, we tend to see that pattern in future observations. As data scientists we have to try and overcome these tendencies.

IMPEDIMENTS TO COLLABORATION

(1/1 point)

What are some impediments to collaboration? Select all the answers that match:

Top of Form

One-off analyses , Diverse tools and programming/scripting languages, Poor script, program, and tool library management, Often easier to rewrite a script instead of searching for it, - correct

One-off analyses Diverse tools and programming/scripting languages Poor script, program, and tool library management Often easier to rewrite a script instead of searching for it

Bottom of Form

**EXPLANATION**

All of the answers are correct. Often, an analysis is exploratory and only performed once. The wide variety of tools and programming and scripting languages makes it difficult to share them, and there are few systems for managing all of the software. Also, the lack of a management system means that it may be easier to reimplement an analysis rather than search for it.

Checkyour answer **Hide Answer**

EXTRACT-TRANSFORM-LOAD

(1/1 point)

Why is Extract-Transform-Load critical? Select all the answers that match:

Top of Form

ETL prepares data for processing, ETL provides the data we need for analyses, - correct

ETL prepares data for processing It is easier to perform data analyses before performing ETLETL provides the data we need for analyses

Bottom of Form

**EXPLANATION**

ETL is the first step in Data Science. We have to acquire and convert data into a canonical format before we can analyze the data. Without ETL, there is no data to analyze.