Test Yourself Questions for Module 1.

Data Scientist Review:

1. Define data science.
2. List four data science skills.
3. Name and briefly distinguish three different types of models discussed in class.
4. Name one thing that separates data science from
   1. machine learning
   2. data mining
   3. statistics
   4. mathematics
   5. database administration
5. What is meant by data cleaning? Give examples.
6. Name 6 components of a data life cycle.
7. Describe a 5-part data science workflow, at each part describe
   1. The main purpose of that part of the life cycle
   2. An issue that could arise to block achievement of the purpose
   3. A strategy you would propose to resolve the block and the mechanism by which it would be resolved
8. What does the FORCE11 acronym FAIR mean? Do you think this initiative would be helpful to science? Justify your answer.
9. Name 3 types of tools data scientists use and what they are useful for
10. What area of data science do you personally find most interesting? Why?