*Your responses can be drawn directly from your Learning Responses. Feel free to copy, paste those answers, and edit as needed.*

1. **From Keith McCormick's LinkedIn Learning presentations**, what **two (2) concepts, insights, or examples** most stood out to you? Briefly summarize it, explain why it stood out to you, and reflect on questions or implications that it raises.
2. **From Siegel's introduction to Predictive Analytics**, report one of the key ideas that most stood out to you, and why.
3. **From your own research**, tell us about the predictive analytics use-case you researched and read about: What was the purpose or problem, who benefited, how much, etc. Also provide your source citation, including a link if relevant.
4. **From Keith McCormick's LinkedIn Learning presentations**, what **two (2) concepts, insights, or examples** most stood out to you?

* I was surprised by the emphasis placed on the very explicit definition presented for data mining. McCormick chose to use the term “data mining” because he uses the cross industries standard process for data mining. They use the term so he does as well. He does this for precision. He lays out the essential elements to define explicitly what is and is not data mining and to draw a distinction between it and other analytics terms like statistical analysis, hypothesis testing, and reporting.
* McCormick says “I’ve never seen a project fail because there weren’t patterns in the data.” I included this one because I found it to be the most surprising of any statement made in the videos. Business patterns lead to data patterns is a logical statement, but McCormick goes further to say that he’s never seen a project fail because the patterns didn’t exist. I would appreciate a dialog about this one. Does this mean that patterns existed but couldn’t be used or possibly that the data miner wasn’t smart enough to recognize or leverage the patterns to create a model? More on this one as I’ll bring it up if we have a group discussion.

1. **A key idea from Siegel: “Geeks Power the World”**

* This was my favorite concept from the introduction of Siegel’s book. It encapsulates the extreme (!!!) fortune that I’ve experienced in my career which is being able to do something so enjoyable while adding outsized value enabling outsized compensation. In other words, I’ve been able to make a lot of money doing fun stuff. Siegel points to the enjoyment as being the defining indicator of geekdom. That’s a good point, but we can’t lose sight of the fact that the opportunity is provided by the value added. Above I discussed the outsized impacts of small improvements resulting from imperfect models. This is the driver of the value add. Software and more specifically models scale well. The marginal cost of applying a model is very nearly zero. Provided the model has even a small positive net value on each application, the resulting upside is only limited by the number of predictions that can be made. Even small things done a bunch of times aggregate to big changes – in this case big profits.

1. **From my research:**

Xu, L., Gholami, S., McCarthy, S., Dilkina, B., Plumptre, A., Tambe, M., Singh, R., Nsubuga, M., Mabonga, J., Driciru, M., Wanyama, F., Rwetsiba, A., Okello, T., & Enyel, E. (2019). *Stay ahead of poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations*. arXiv. <https://arxiv.org/abs/1903.06669>

The authors developed enhancements to the Protection Assistant for Wildlife Security (PAWS), which is an ML pipeline that predicts areas at high risk for poaching activity. Their enhancements mitigated data quality issues and improved detection of snares by 30%. PAWS generally is used to optimize patrol routes. Beneficiaries include wildlife management officials as well as obviously the animals themselves.

When looking for a response to this question I intentionally sought out something obscure. These models were created using data from Uganda and Cambodia then applied in the same regions. It would be stereotyping to assume that the data from game managers in developing countries is bad, but that is indeed the case. The managers concerns are life or death for both them and the wildlife. One focus of the study is how to deal with the “unreliable and imbalanced data”. An example, 99.6% of the labels were negative indicating no illegal activity. Was there truly no activity, or was it just well concealed? The fact that the authors were able to have a positive impact and get quality results from their models is encouraging and inspiring.