**Unsupervised learning capstone**

For your third capstone, you'll complete an unsupervised learning project. You can choose your own research question or choose from one below. In this capstone, you will be graded by your peers.

How can you use clustering techniques to achieve the following?

1. Netflix wants to identify similar movies based on movie characteristics.
2. A company wants to segment its customers.
3. Your startup is creating an innovative new email software. How can you help them create an effective spam filter?
4. Your chief marketing officer wants to know how much website traffic is due to spam and bots versus actual customers. How can you group the traffic?
5. Your spouse loves fantasy baseball and has a few favorite players. How can you help them to choose who to have on their team?
6. The security team in a bank wants to identify fraudulent transactional activities.

Here are the next steps:

1. **Go out and find a dataset of interest.** It could be one that helps you work on one of Thinkful's recommended research questions, or it could be any other dataset that addresses an unsupervised learning question of your own.
2. **Explore the data.** Get to know the data. Spend a lot of time going over its quirks. You should understand how it was gathered, what's in it, and what the variables look like.
3. **Try several different approaches.** Really work to tune a variety of models before choosing what you consider to be the best performer.

Keep the following considerations in mind: How do clustering and modeling compare? What are the advantages of each? Why would you want to use one over the other?

This will ultimately include the following deliverables:

* A Jupyter Notebook that tells a compelling story about your data. You'll submit this Notebook at the end of this checkpoint.
* A 15-to-30-minute presentation of your findings. You'll need to produce a deck and present it to your peers.

Conduct the analysis in Jupyter. Provide a complete research report, using the research presentation framework introduced in the *Presenting research results* checkpoint as a starting point. The report should use compelling visualizations and actionable insights to tell the story to your intended audience. Walk through the analysis using clean, reproducible code. Include plenty of notes and comments to guide others through your thinking.

Rather than create a deck in a presentation tool like PowerPoint or Google Slides, use Jupyter's built-in slide deck capabilities. This cuts down on the time-consuming and error-prone process of copying and pasting into an external program. It will also make your presentation fully reproducible.

On the day of your presentation, you'll give your presentation to your peers. Members of the audience will participate in asking you questions, and you will ultimately be graded by your peers according to the following rubric.

* **The presenter delivered the material in a clear and structured manner.** The presentation had a great structure to deliver the results of the analysis with all parts aiding in the goals of the presentation. No tangents that detracted from the story. The presenter seemed very prepared and exhibited great presentation skills.
* **The presenter maintained my interest during the entire presentation.** It was very interesting and the presenter held my attention the entire time. The presenter interacted with the audience to make sure they followed along.
* **The presenter answered questions effectively.** The presenter was able to provide great answers to all questions. The answers all considered the technical level of the audience and had good interpretations.
* **The presenter clearly presented the conclusion and recommendations of the research study.** The presenter very clearly described how *k* was determined (silhouette versus elbow versis Davies-Bouldin), or how it might support a business objective. The presenter was highly effective at discussing logical next steps and recommendations for how the model can be used as-is, and what things would help improve model accuracy.
* **The presentation was concise and informative.** The presenter was able to present the challenges, the process, and the conclusions in a brief, logical, and comprehensive manner, enabling the audience to gain an understanding of the material presented. The presenter provided an accurate and sufficiently complete explanation of key concepts.
* **The presentation provided sufficient details about methods to understand how the study was conducted but wasn't too detailed.** The presenter was highly effective at conveying the methods used throughout the study, without going into more detail than necessary. The presenter was highly effective at explaining how they selected a model, and how the model was appropriate for the business need. The presenter was highly effective at discussing feature importance and identifying extreme features.
* **The presentation contained practical examples and useful techniques that applied to current work.** The presenter was highly effective at demonstrating appropriateness of the model used. The presenter was highly effective at identifying which mistakes or assumptions were the most costly.
* **The visual aids were effective.** The presenter was highly effective in using graphics and visual aids that were clean, easy to follow, and told a compelling story without excess text or complex charts. The presenter was highly effective in using visual aids to support their story, instead of reading from the slides word-by-word.

Remember to focus on desirable presentation skills such as an appropriate pace, the right depth of coverage for the audience, and effective use of examples and visuals.

Be prepared to talk about the following:

* Your end user and the value that your project provides to them
* Quantifiable results based on the research that you've conducted
* The visuals that you chose and what they represent for your research
* Next steps that you'd take to expand on your work
* A specified research question that your model addresses
* How you chose your model specification and what alternatives you compared it to
* The practical uses of your model for an audience of interest
* Any weak points or shortcomings of your model