1. What software did you use for the clustering?

I primarily use Python to conduct my research as it is open source and freely available. There is a fantastic scientific ecosystem within the Python community and so there are a number of well-made libraries for clustering. I prefer a Scikit-Learn as it interacts nicely with the rest of a collection of libraries known as the scientific stack (NumPy, pandas, etc.).

1. What clustering approaches have you tried out?

I have tried a number of the established algorithms including k-means, DBSCAN and hierarchical methods. However, I have found that since I am dealing with large-scale data, k-means suits my domain well as it is easily scalable. More importantly is what you decide to cluster on. Clustering raw data can be incredibly useful but, as with all methods, the inclusion of prior expert knowledge is powerful. So, clustering patients based on their use patterns rather than clustering the episodes is more useful for learning about the kinds of patients in the system, and their behaviours.

1. Where are you taking your work next?

I aim to utilise clustering techniques to identify areas of potential improvement or current “success” within the system. From there, further statistical analysis can be done giving more meaningful results. The hope beyond that is to build game- and queue-theoretic models of patients and service providers to understand their behaviours and interactions at a lower level.