Q - Let's say we work at a credit card company and we're trying to figure out if people are going to pay their bills on time. We have everyone's purchases, split into four main categories: groceries, dining out, utilities, and entertainment. What are some ways you might use KNN to create this model? What aspects of KNN would be useful?

A – The first major thing I would look at is how each of the four categories relates to likelihood of paying a bill on time. This is for two reasons, the first being simply to try and understand the data a bit better and the second is to see if there is a possibility for PCA. If for example Groceries and Dining out has the same variance and people tend to pay their bills on time at the same pace for both it might be worth combining them.

I’m also not entirely sure what a 4-D plot would look like(I know it’s not actually possible to plot it but theoretically I’m just trying to imagine). With that in mind I might try to create several 2-D plots with various pairs of purchase types.

Ultimately I need to know more about what question is being asked. Above it says it wants to know if ‘people will pay their bills on time’ and then says we have purchase data, am I going off the assumption that we think purchases might affect that? As in people whose majority purchases are on one or two purchase types vs another are more likely to pay bills on time? If so then my original thinking is correct.

One thing that would need to be done is to standardize the data so all purchase types are on the same. This needs to be done because purchases of things like entertainment can vary widely from month to month whereas groceries might be stable, and utilities are more seasonal.