## Accuracy may not always be the best metric to use

### Detecting credit card fraud

We have a bunch of data in the form of credit card transactions. There are 284,335 good transactions and 472 fraudulent transactions. This is actually real data. So, let's try to come up with a model that has great accuracy..in other words, a model that is correct most of the time.

Can you think of a model that has over 99 percent accuracy?

### All the transactions are good

Let's use a model that says, all the transactions are good. This model has over 99 percent accuracy, which means it is correct over 99 percent of the time. Now, what is the accuracy of this model, namely how many times are we correct? Well, we're correct 284,335 times out of 384,887 times. And that is 99.83 percent. Logically, this model must be pretty good if it's accurate is that high, right? No! This model is not catching any of the bad ones. And the point of the model is to catch the fraudulent transactions.

### All transactions fraudulent

Can we get a model that catches all the bad transactions? This model catches all fraudulent transactions. Is that a good model? No! That's also a terrible model since it's accidentally catching all the good ones.

It's pretty tricky to just look at accuracy and use that to evaluate our model because it may completely miss the point when the data is skewed like this one.

### **Quiz Question**

Accuracy is the answer to the question

1. "Out of all the transactions, how many did we classify correctly?"
2. "Out of all the good transactions, how many did we classify correctly?"
3. "Out of all the fraudulent transactions, how many did we classify correctly?"