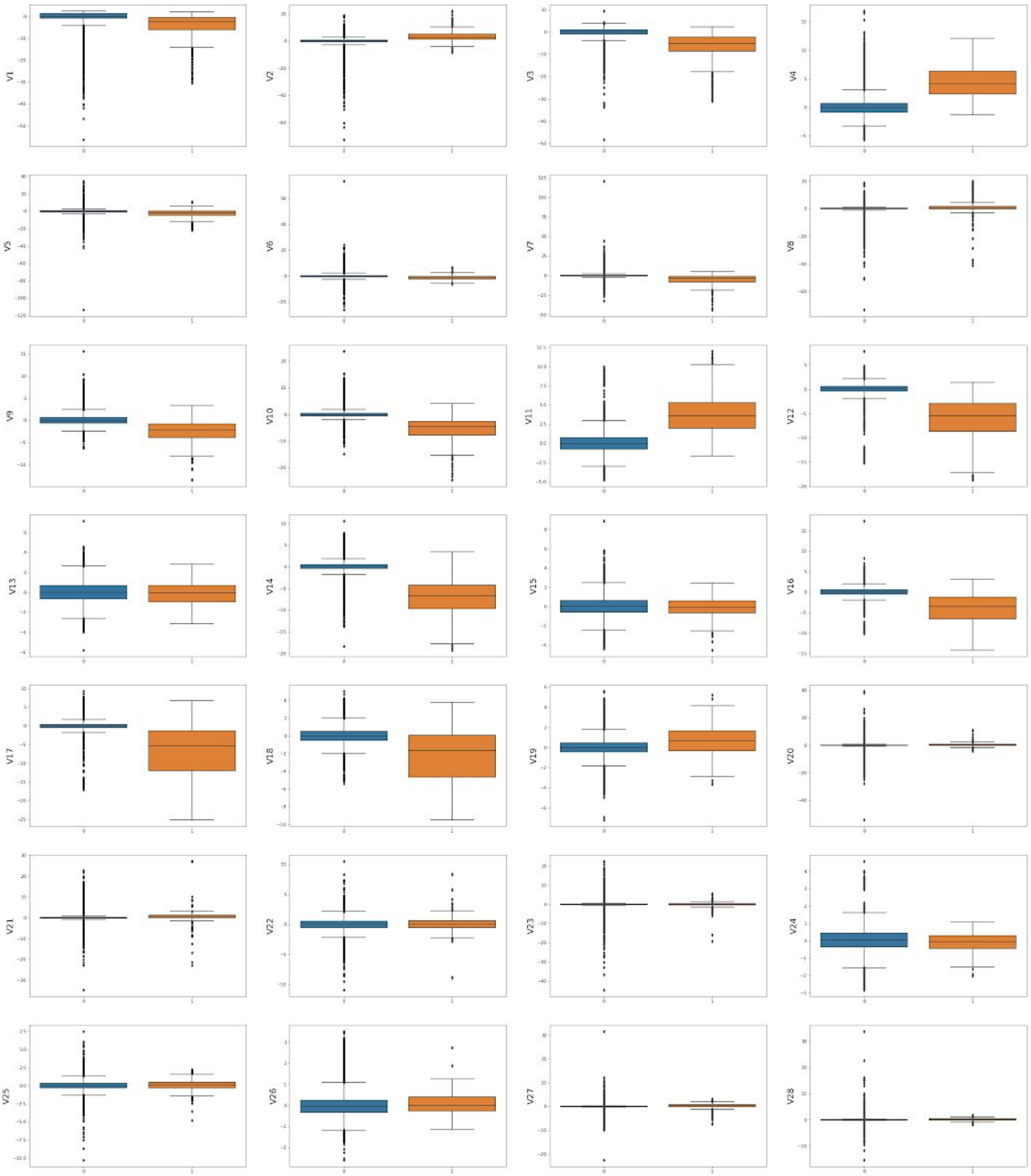

This plot gives us a good sense of what features might play a significant role in determining whether a transaction is fraud or not. For example, for features including (V4, V11,V12, V14, V16, V17), the shape and position of fraudulent(blue) and valid (orange) boxes are very different, which means this feature of valid and fraudulent transactions are very different in distribution. Based on this, we now have a sense that these features are important in classifying valid and fraudulent transaction. By contrast, feature such as (V6, V8, V13, V15, V21, V23, V25, V27, V28), their valid and transactions have similar distribution of this feature, so they might be irrelevant factors. These indications will be further supported by a quick random forest, which is shown in the latter part of this notebook (see my full code at: https://github.com/MengSunS/Credit-Card-Fraud-Detection-/blob/master/CreditCardFraudDetection.ipynb).