# Data Question 1

**Objectives:**

- How many distinct phone numbers had at least 2 sim swap event changes in this data?
- How many records had at least 3 change events?
- What questions do you have about the data?
- **File Name: dataquestion1.csv**
- **Please return your code that answers these questions.**

# Data Question 2

You created a model to predict the likelihood that a consumer application was fraudulent at the time of onboarding. 1 = fraudulent transaction. You are trying to evaluate the model with this confusion matrix.

1. What does the confusion matrix tell you about the data?
2. What questions do you have about the problem/data?
3. Anything you would do differently to measure the model?
4. How could you improve the model?
5. If you have an example/project where the data was imbalanced, send us the link to the code and briefly (<3 paragraphs) describe the project and issues you dealt with and how you resolved them.



Confusion matrix

|  | Predicted label 0 | Predicted label 1 |
|---|---|---|
| True label 0 | 3000 | 1000 |
| True label 1 | 25 | 75 |