**Homework #2**

**Due: Sunday, October 14, 2018**

**Note:** Submit your homework by pushing your code to your Gitlab account. Please use Google Colab as mentioned during the lecture.

1. **REQUIRED PART**

**Goal**

E-commerce websites often transact huge amounts of money. And whenever a huge amount of money is moved, there is a high risk of users performing fraudulent activities, e.g. using stolen credit cards, doing money laundry, etc.

Machine Learning really excels at identifying fraudulent activities. Any website where you put your credit card information has a risk team in charge of avoiding frauds via machine learning.

The goal of this challenge is to build a machine learning model that predicts the probability that the first transaction of a new user is fraudulent.

**Description**

Company XYZ is an e-commerce site that sells hand-made clothes. You have to build a model that predicts whether a user has a high probability of using the site to perform some illegal activity or not. This is a super common task for data scientists.

You only have information about the user first transaction on the site and based on that you have to make your classification ("fraud/no fraud").

These are the tasks you are asked to do:

For each user, determine her country based on the numeric IP address.

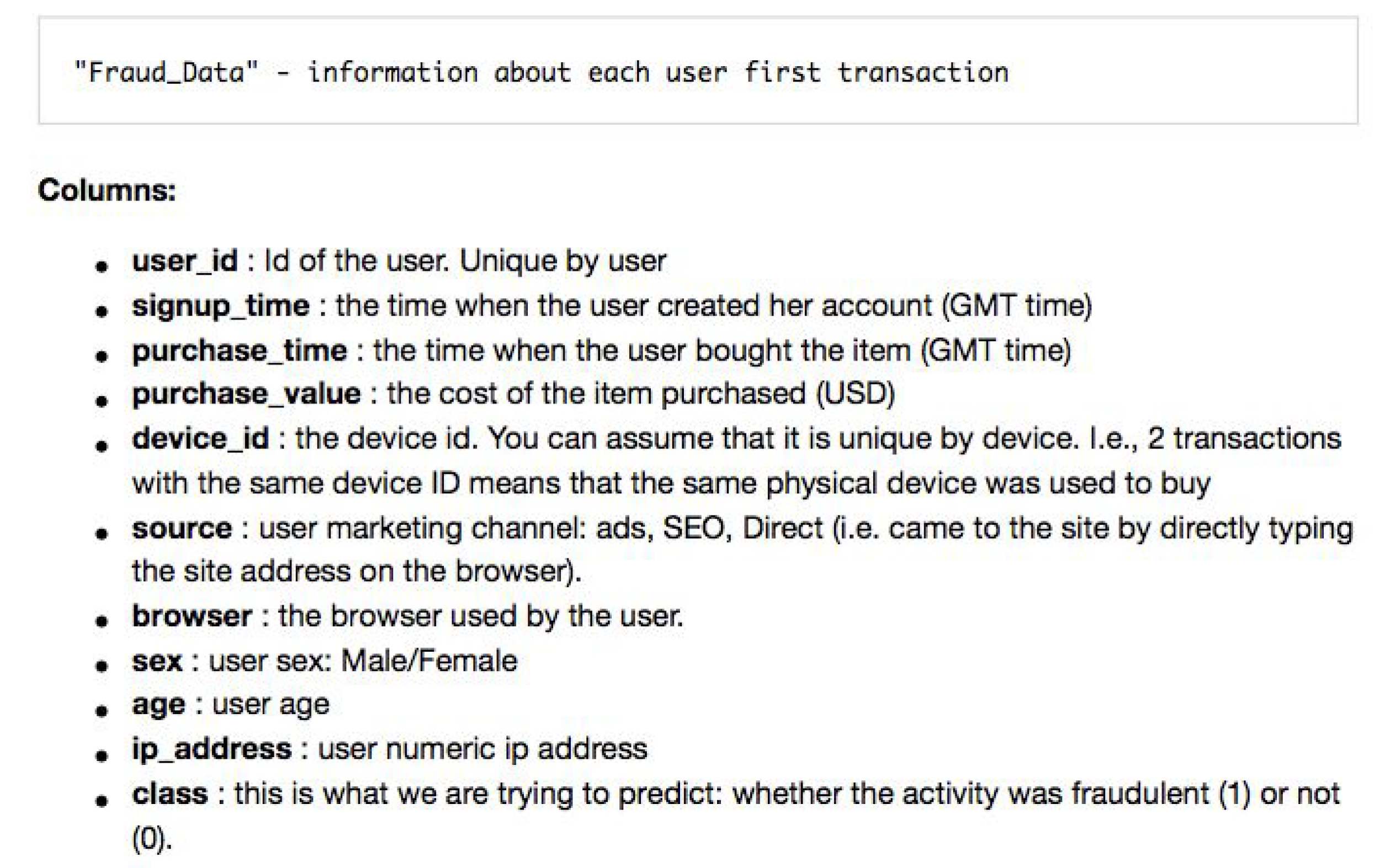
Build a model to predict whether an activity is fraudulent or not. Explain how different assumptions about the cost of false positives vs false negatives would impact the model.

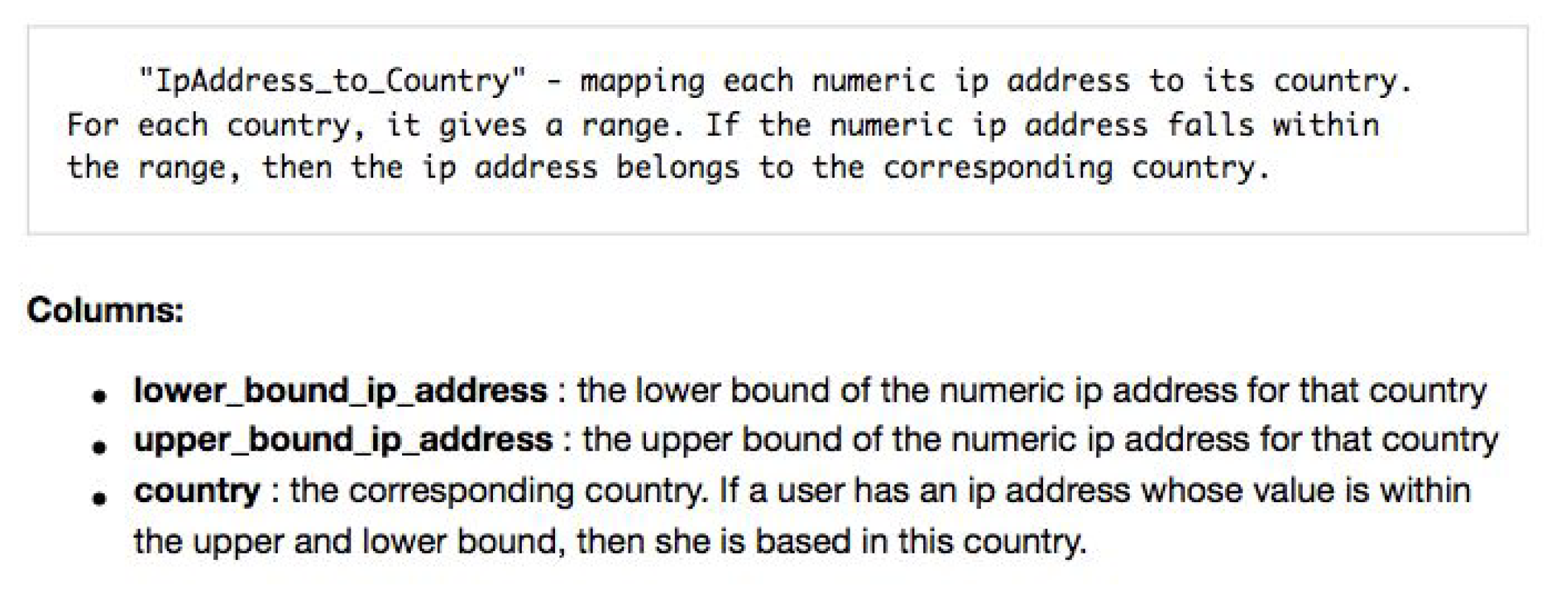
Your boss is a bit worried about using a model she doesn't understand for something as important as fraud detection. How would you explain her how the model is making the predictions? Not from a mathematical perspective (she couldn't care less about that), but from a user perspective. What kinds of users are more likely to be classified as at risk? What are their characteristics?

Let's say you now have this model which can be used live to predict in real time if an activity is fraudulent or not. From a product perspective, how would you use it? That is, what kind of different user experiences would you build based on the model output?

**Data Information**

Please check the links of data below:





**Dataset Link**

<https://drive.google.com/file/d/1TwtrXiRHVFbrga061gux7qvGV5l-tDzL/view?usp=sharing>

<https://drive.google.com/file/d/15uWrFKmYVXsF99xj1B-4sDBAIlCYV8zr/view?usp=sharing>