**ATD Data Science Assessment**

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

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.

Electronica is an e-commerce site that sells wholesale electronics. You have been contracted to build a model that predicts whether a given transaction is fraudulent or not. You only have information about each user’s first transaction on Electronica’s website. If you fail to identify a fraudulent transaction, Electronica loses money equivalent to the price of the fraudulently purchased product. If you incorrectly flag a real transaction as fraudulent, it inconveniences the Electronica customers whose valid transactions are flagged—a cost your client values at $8.

**Instructions:**

- explore the data

- keep all your work

- comment your code

- explain any features that you engineer

- indicate which model is final

- originality / exploration of different model types

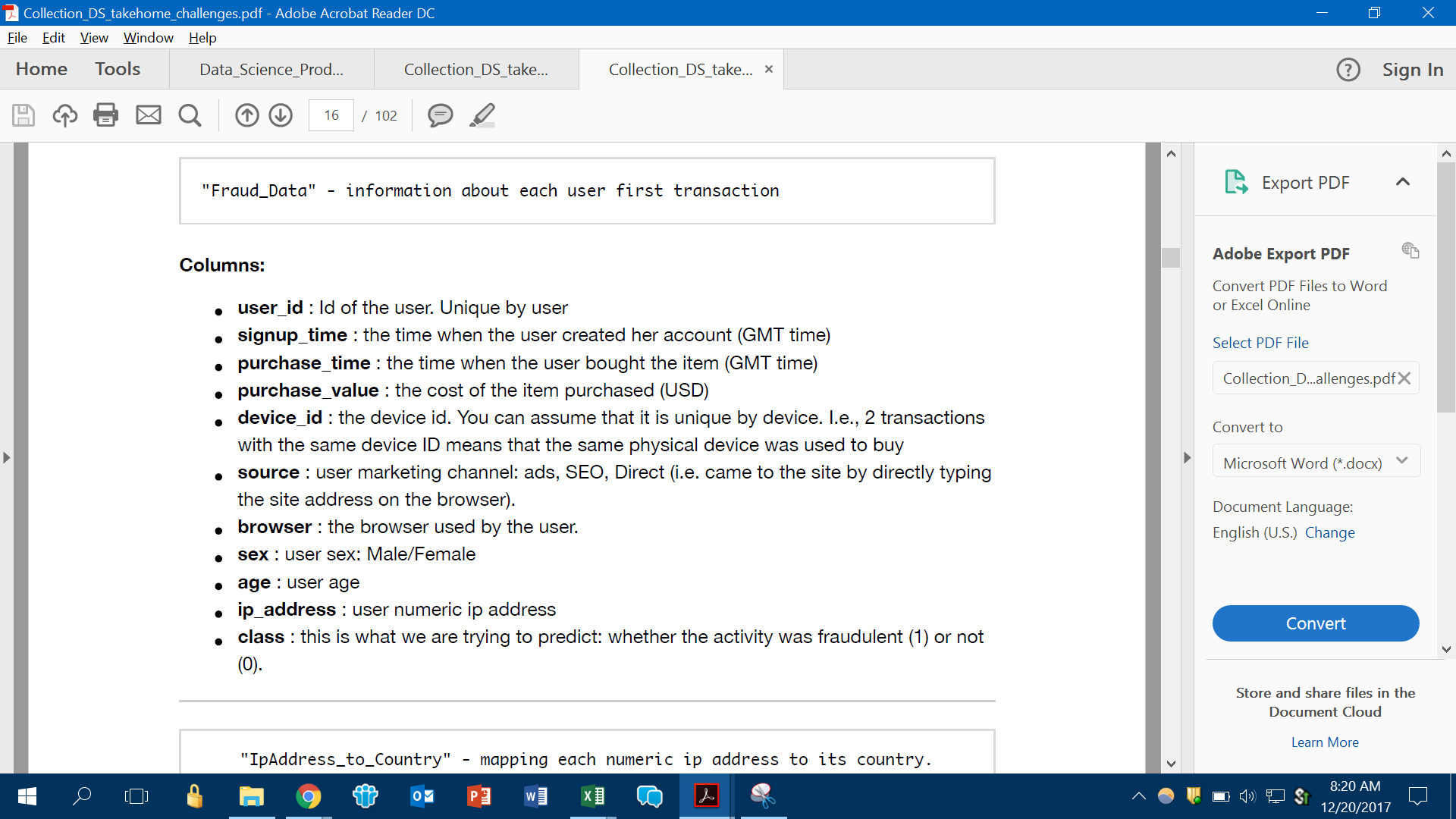
- feel free to include “next steps” in the comments of your code

- feel free to use whatever language/method you are comfortable with (python, R, julia, etc.); I would recommend using a jupyter notebook though for easy visualization

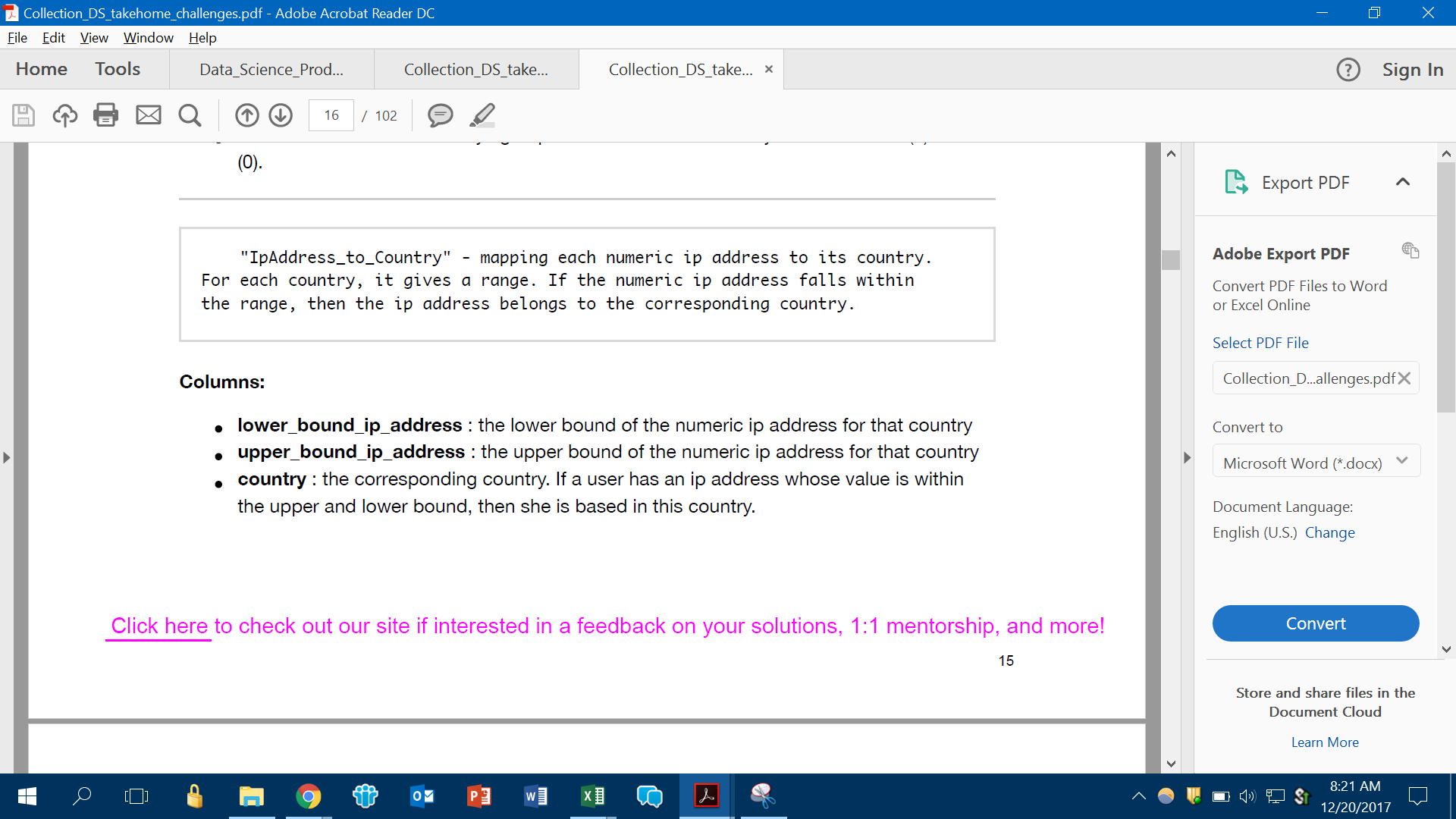
- PLEASE don’t spend a ridiculous amount of time trying to squeeze the last drop of performance from your models; just make sure your code is well thought-out and you can justify any choices you make

**(more on the next page)**

**Fraud data**

****

**IP address to country mapping**

****