- **Business problem:**

Synthetic identity fraud on credit card applications. It occurs when a credit card applicant creates a fake identity using a combination of real and fake information, such as a real social security number and a fake name or address.

- **What events/things the algorithm will score for possible fraud:**

Credit card applications. Assign a numerical probability that the application is a synthetic identity fraud.

- **Likely data and fields:** The algorithms will score each application. We'll focus on the Identity fields on the application. Name, date of birth, ssn, zipcode, address, phone number, and other identifying information such as ip address and device id. Also, we can look at credit history and credit/loan amount. we can use past known fraud examples to build supervised learning models. Other auxiliary data can include other identity data
    1. Census Bureau data (zipcode)
    2. Employment data (we can check if the employement information is consistent with the income)
    3. Phone book data (Besides from the name, address, phone number, we can check if the phone number is virtual or disposable)
    4. Credit Bureaus
    5. Known fraudulent or risky locations, people, email, phone number

- **What to look for:** Indication of synthetic identity fraud.
  The possibility of fraud goes up when the following occurs:

1. There are many applications coming from the same address/phone number/ssn/ip address or the same combination groups
2. The same entity or combination group showed up in abnormal frequency (for instance, the application with the same combination of ssn and phone number applied six times a week
3. The relative velocity, which is ratio of the short-term velocity to a longer-term averaged velocity, is high
4. The day of week risk is high
5. The unique number of an entity or a combination for a particular entity/group are high
6. High credit limits or loan amounts that seem unusual or disproportionate to the individual's known income or employment