**Dataset Summary**

The dataset consists of five files :

1. test\_identity.csv
   1. Identity Dataset to test performance of model
2. test\_transaction.csv
   1. Transaction Dataset to test performance of model
3. train\_identity.csv
   1. Identity Dataset to be used for model training
4. train\_transaction.csv
   1. Transaction Dataset to be used for model training
5. sample\_submission.csv
   1. File containing prediction results

Column names in train and test identity dataset has different formats. We renamed test identity columns name to match to that of train column names.

**Transaction Dataset :**

Columns in the transaction dataset can be grouped as below:

1. TransactionID
   1. ID of the transaction and is used to connect to identity dataset
2. isFraud
   1. Indicating if the transaction is fraud or not
3. TransactionAmt
   1. Transaction amount (in USD)
4. ProductCD
   1. Product code with 5 distinct categories
   2. Categorical Features
5. card1 - card6
   1. Payment card information
   2. Card 4 contains 4 distinct categories of the card used for the transaction
   3. Card5 has 120 unique values
   4. Card6 has 4 distinct categories
   5. Categorical Features
   6. Actual values are masked for some columns
6. addr1 - addr2
   1. addr1 - billing region
   2. addr2 - billing country
   3. Categorical Features
7. dist1 - dist2
   1. Distance
   2. Categorical Features
8. P\_emaildomain
   1. purchaser email domains
   2. Categorical Features
9. R\_emaildomain
   1. recipient email domains
   2. Categorical Features
10. C1 - C14
    1. counting, such as how many addresses are found to be associated with the payment card, etc.
    2. The actual meaning is masked
11. D1 - D15
    1. timedelta, such as days between previous transaction, etc
12. M1 - M9
    1. match, such as names on card and address, etc.
    2. Categorical Features
13. V1 - V339
    1. Vesta engineered rich features, including ranking, counting, and other entity relations.

**Identity Dataset :**

1. TransactionID
   1. ID of the transaction and is used to connect to transaction dataset
2. DeviceType
   1. Device type used for transaction mobile or desktop
   2. Categorical Features
3. DeviceInfo
   1. Information about the device used for transaction
   2. Categorical Features
4. id-01 to id-38
   1. Masked information about the device