apply

October 7, 2024

[3]:

```
import pandas as pd
```

0.0.1 Using apply

Allows you to pass a function and apply it on every value of the Pandas series.

- create a custom function() issuer, similar to the functions practice.
- apply() theissuer() function to each row to generate a new column “issuer”

[4]:

```
def issuer(card: str) -> str:
    card = str(card)
    first_digit = card[0]
    if first_digit == "6":
        return "DISC"
    if first_digit == "5":
        return "MC"
    if first_digit == "4":
        return "VISA"
    if first_digit == "3":
        return "AMEX"

    return "Unknown"


assert issuer("5295474999519325") == "MC"
assert issuer("4584857473589512") == "VISA"
assert issuer("6011687948781644") == "DISC"
assert issuer(364733448466728) == "AMEX"
```

Bringing in a text file …

- We are reading a text file - credit_cards.txt using pd.read_csv.
- sep=” ” specifies that the columns in the file are separated by spaces.
- header=None tells us that the file doesn’t have a header row, so column names will be added later.
- df.columns = [“card”, “exp”] - we are adding two columns named “card” and “exp”. “card”

```
[5]: df = pd.read_csv("https://raw.githubusercontent.com/mafudge/datasets/master/
      ↪credit-cards/credit_cards.txt", sep=" ", header=None)
     df.columns = ["card", "exp"]
     df.head()
```

```
[5]:                card      exp
     0  6011325926714465  02/2023
     1  6011687948781644  02/2023
     2  4567958456643465  07/2022
     3  4235823774487478  03/2023
     4  4982444199283999  12/2022
```

Adding the function

Add a new column called 'issuer':

- Code below applies a function issuer to each row of the DataFrame df, where the function issuer takes the value from the "card" column as an argument.
- The result of this function is then stored in a new column 'issuer'.
- axis=1 ensures that the function is applied row-wise (i.e., it operates on each row instead of columns).

```
[6]: df['issuer'] = df.apply(lambda row: issuer(row["card"]), axis=1)
     df.head()
```

```
[6]:                card      exp issuer
     0  6011325926714465  02/2023   DISC
     1  6011687948781644  02/2023   DISC
     2  4567958456643465  07/2022   VISA
     3  4235823774487478  03/2023   VISA
     4  4982444199283999  12/2022   VISA
```

```
[7]: df.issuer.value_counts()
```

```
[7]: issuer
     MC      3863
     VISA    3645
     AMEX    1496
     DISC     996
     Name: count, dtype: int64
```

```
[18]: df.issuer.value_counts(normalize=True)
```

```
[18]: issuer
      MC      0.3863
      VISA    0.3645
      AMEX    0.1496
```

```
DISC    0.0996
Name: proportion, dtype: float64
```

[ ]:

```

```