

Chapter 7 - Exercise 5: Trực quan hóa dữ liệu Loan

Cho tập dữ liệu `loan_train.csv`

Tập dữ liệu gồm có 346 khách hàng, KH đã trả hết các khoản vay hoặc còn đang trả. Giải thích các thuộc tính sau:

Field	Description
Loan_status	Whether a loan is paid off on in collection (khoản vay đã trả hết hoặc còn đang trả)
Principal	Basic principal loan amount at the (Số tiền cho vay gốc)
Terms	Origination terms which can be weekly (7 days), biweekly, and monthly payoff schedule (lịch trình thanh toán hàng tháng)
Effective_date	When the loan got originated and took effects (Ngày thực_hiện khi khoản vay có nguồn gốc và có hiệu lực)
Due_date	Since it's one-time payoff schedule, each loan has one single due date (Ngày đến hạn)
Age	Age of applicant (Tuổi của KH vay)
Education	Education of applicant (Trình độ học vấn)
Gender	The gender of applicant (Giới tính)

```
In [1]: 1 import numpy as np
        2 import pandas as pd
        3 import matplotlib.pyplot as plt
        4 import seaborn as sns
```

```
In [2]: 1 loan = pd.read_csv('data/loan_train.csv')
        2 loan.drop(columns=['Unnamed: 0', 'Unnamed: 0.1'], inplace=True)
```

```
In [3]: 1 loan.shape
```

Out[3]: (346, 8)

```
In [4]: 1 loan.head()
```

```
Out[4]:
```

	loan_status	Principal	terms	effective_date	due_date	age	education	Gender
0	PAIDOFF	1000	30	9/8/2016	10/7/2016	45	High School or Below	male
1	PAIDOFF	1000	30	9/8/2016	10/7/2016	33	Bechalor	female
2	PAIDOFF	1000	15	9/8/2016	9/22/2016	27	college	male
3	PAIDOFF	1000	30	9/9/2016	10/8/2016	28	college	female
4	PAIDOFF	1000	30	9/9/2016	10/8/2016	29	college	male

In [5]:

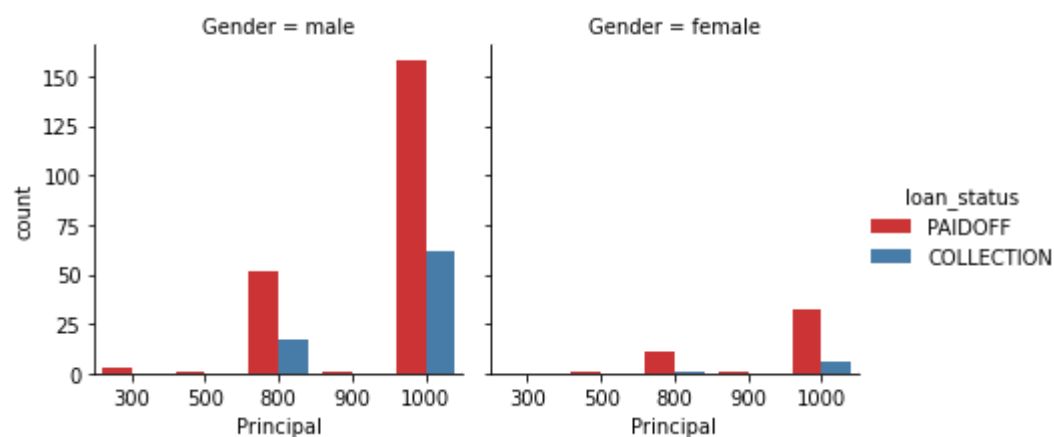
```
1 loan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 346 entries, 0 to 345
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   loan_status           346 non-null   object 
1   Principal              346 non-null   int64  
2   terms                 346 non-null   int64  
3   effective_date        346 non-null   object 
4   due_date              346 non-null   object 
5   age                   346 non-null   int64  
6   education             346 non-null   object 
7   Gender                346 non-null   object 
dtypes: int64(3), object(5)
memory usage: 21.8+ KB
```

In [6]:

```
1 # Câu 1: Vẽ biểu đồ biểu diễn tần suất của Principal theo hai nhóm: Gender và Loan_
2 # cho nhận xét
3
```

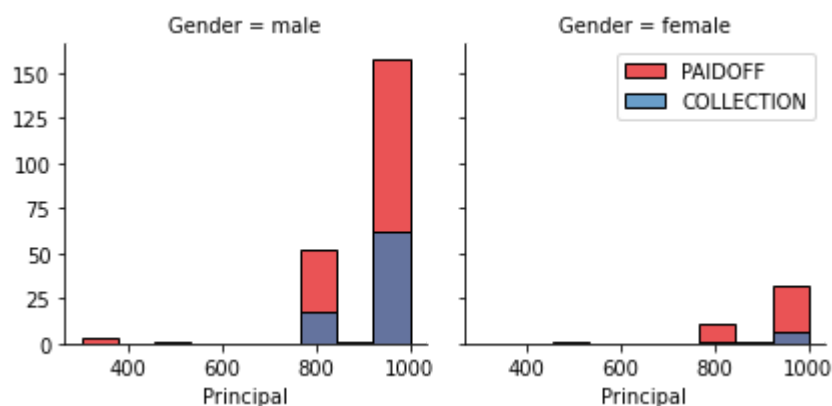
Nhấn vào đây để xem kết quả!



In [7]:

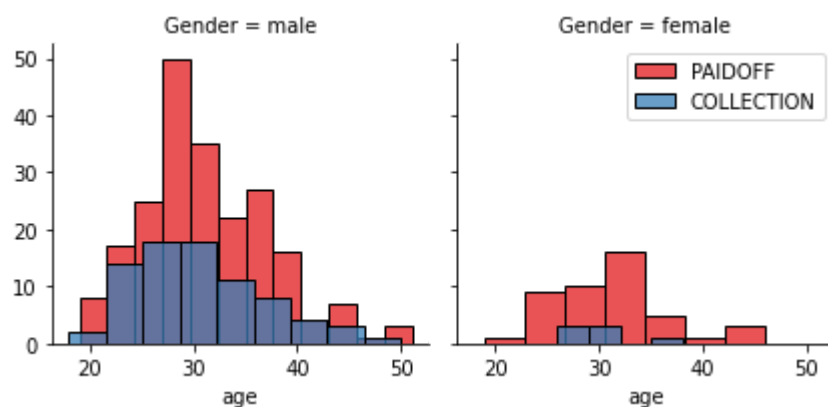
```
1 # Câu 2: Vẽ biểu đồ biểu diễn phân phối của Principal theo hai nhóm: Gender và Loan_
2 # cho nhận xét
3
```

Nhấn vào đây để xem kết quả!



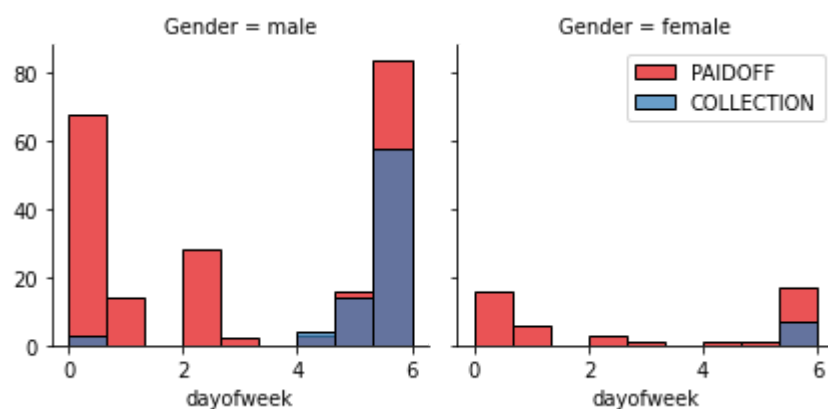
```
In [8]: 1 # Câu 3: Vẽ biểu đồ biểu diễn phân phối của age theo hai nhóm: Gender và Loan_status
2 # cho nhận xét
3
```

Nhấn vào đây để xem kết quả!



```
In [9]: 1 # Câu 4: Vẽ biểu đồ biểu diễn phân phối của dayofweek theo hai nhóm: Gender và Loan_status
2
3 # tạo cột dayofweek (Monday=0, Sunday=6)
4 loan['dayofweek'] = pd.to_datetime(loan.effective_date).dt.dayofweek
5
6 # vẽ biểu đồ
7
```

Nhấn vào đây để xem kết quả!



```
In [ ]: 1
```