

Import thư viện

```
In [1]: import pandas as pd
```

1. Đọc file dữ liệu

```
In [2]: thietbi = pd.read_csv('../data/thietbi.csv')
nguoidung_thietbi = pd.read_csv('../data/nguoidung_thietbi.csv')
nguoidung_sudung = pd.read_csv('../data/nguoidung_sudung.csv')
```

2. Vẽ mô hình quan hệ (ERD) từ bảng dữ liệu trên

```
In [ ]:
```

3. Đổi tên các cột dữ liệu cần thiết

```
In [3]: thietbi.rename(columns={'Retail Branding': 'Branding', 'Marketing Name': 'MarketingName'}, inplace=True)
```

```
In [4]: thietbi
```

```
Out[4]:
```

| | Branding | MarketingName | Device | Model |
|-------|------------|---------------|-------------|---------------------------|
| 0 | NaN | NaN | AD681H | Smartfren Andromax AD681H |
| 1 | NaN | NaN | FJL21 | FJL21 |
| 2 | NaN | NaN | T31 | Panasonic T31 |
| 3 | NaN | NaN | hws7721g | MediaPad 7 Youth 2 |
| 4 | 3Q | OC1020A | OC1020A | OC1020A |
| ... | ... | ... | ... | ... |
| 14541 | pendo | PNDPP44QC10 | PNDPP44QC10 | PNDPP44QC10 |
| 14542 | pendo | PNDPP44QC7 | PNDPP44QC7 | PNDPP44QC7 |
| 14543 | sugar_aums | QPOINT | QPI-1 | QPI-1 |
| 14544 | tecmobile | OmnisOne | OmnisOne | Omnis One |
| 14545 | ucall | EASY1 | EASY1 | EASY1 |

14546 rows × 4 columns

4. Trộn bảng nguoidung_sudung và nguoidung_thietbi thông qua inner join để được bảng kết quả có các cột sau: [outgoing_mins_per_month, outgoing_sms_per_month, monthly_mb, use_id, platform, device]

```
In [5]: dfKetQua = pd.merge(nguoidung_sudung,
                             nguoidung_thietbi[['use_id', 'platform', 'device']],
                             on='use_id')
```

```
In [6]: dfKetQua
```

```
Out[6]:
```

| | outgoing_mins_per_month | outgoing_sms_per_month | monthly_mb | use_id | platform | device |
|-----|-------------------------|------------------------|------------|--------|----------|--------------------|
| 0 | 21.97 | 4.82 | 1557.33 | 22787 | android | GT-I950 |
| 1 | 1710.08 | 136.88 | 7267.55 | 22788 | android | SM G930 |
| 2 | 1710.08 | 136.88 | 7267.55 | 22789 | android | SM G930 |
| 3 | 94.46 | 35.17 | 519.12 | 22790 | android | D230 |
| 4 | 71.59 | 79.26 | 1557.33 | 22792 | android | SM G361 |
| ... | ... | ... | ... | ... | ... | . |
| 154 | 198.59 | 90.49 | 5191.12 | 23043 | android | SM G900 |
| 155 | 198.59 | 90.49 | 3114.67 | 23044 | android | SM G900 |
| 156 | 106.65 | 82.13 | 5191.12 | 23046 | android | Moto (4 |
| 157 | 344.53 | 20.53 | 519.12 | 23049 | android | SM G900 |
| 158 | 42.75 | 46.83 | 5191.12 | 23053 | android | Vodafone Sma ultra |

159 rows × 6 columns



Hoặc dùng left-join tùy theo nhu cầu

```
In [7]: # hoặc left-join
dfKetQua = pd.merge(nguoidung_sudung,
                     nguoidung_thietbi[['use_id', 'platform', 'device']],
                     on='use_id', how='left')
```

In [8]: dfKetQua

Out[8]:

| | outgoing_mins_per_month | outgoing_sms_per_month | monthly_mb | use_id | platform | device |
|-----|-------------------------|------------------------|------------|--------|----------|----------|
| 0 | 21.97 | 4.82 | 1557.33 | 22787 | android | GT-I9505 |
| 1 | 1710.08 | 136.88 | 7267.55 | 22788 | android | SM-G930F |
| 2 | 1710.08 | 136.88 | 7267.55 | 22789 | android | SM-G930F |
| 3 | 94.46 | 35.17 | 519.12 | 22790 | android | D2303 |
| 4 | 71.59 | 79.26 | 1557.33 | 22792 | android | SM-G361F |
| ... | ... | ... | ... | ... | ... | ... |
| 235 | 260.66 | 68.44 | 896.96 | 25008 | NaN | NaN |
| 236 | 97.12 | 36.50 | 2815.00 | 25040 | NaN | NaN |
| 237 | 355.93 | 12.37 | 6828.09 | 25046 | NaN | NaN |
| 238 | 632.06 | 120.46 | 1453.16 | 25058 | NaN | NaN |
| 239 | 488.70 | 906.92 | 3089.85 | 25220 | NaN | NaN |

240 rows × 6 columns

5. Tiếp tục inner-join bảng thietbi và dfKetQua để có thêm 2 cột Model và Branding trong bảng kết quả cuối cùng. Lưu ý ở đây khóa liên kết sẽ khác tên

In [9]: *# inner-join*

```
dfKetQua = pd.merge(dfKetQua,
                    thietbi[['Branding', 'Model']],
                    left_on='device',
                    right_on='Model')
```

In [10]: dfKetQua

Out[10]:

| | outgoing_mins_per_month | outgoing_sms_per_month | monthly_mb | use_id | platform | device |
|-----|-------------------------|------------------------|------------|--------|----------|--------------------|
| 0 | 21.97 | 4.82 | 1557.33 | 22787 | android | GT-I950 |
| 1 | 69.80 | 14.70 | 25955.55 | 22801 | android | GT-I950 |
| 2 | 249.26 | 253.22 | 1557.33 | 22875 | android | GT-I950 |
| 3 | 249.26 | 253.22 | 1557.33 | 22876 | android | GT-I950 |
| 4 | 83.46 | 114.06 | 3114.67 | 22880 | android | GT-I950 |
| ... | ... | ... | ... | ... | ... | . |
| 196 | 22.85 | 34.54 | 6577.12 | 23029 | android | HTC One_M |
| 197 | 180.18 | 17.49 | 2076.45 | 23039 | android | SM G531 |
| 198 | 12.85 | 58.32 | 74.40 | 23040 | android | HTC Desir 62 |
| 199 | 42.75 | 46.83 | 5191.12 | 23053 | android | Vodafone Sma ultra |
| 200 | 42.75 | 46.83 | 5191.12 | 23053 | android | Vodafone Sma ultra |

201 rows × 8 columns



Hoặc left-join

```
In [11]: # left-join
# inner-join
dfKetQua = pd.merge(dfKetQua,
                    thietbi[['Branding', 'Model']],
                    left_on='device',
                    right_on='Model', how='left')
```

6. Hãy liệt kê 5 dòng đầu tiên các mẫu thiết bị (cột device) bắt đầu bằng GT