

```
In [145]: import numpy as np
import pandas as pd
pd.set_option('display.max.columns', 100)
pd.set_option('display.max_rows', 500)
import pandasql as pds
# to draw pictures in jupyter notebook
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
# we don't like warnings
# you can comment the following 2 lines if you'd like to
import warnings
warnings.filterwarnings('ignore')
```

```
In [146]: user_usage = pd.read_csv("data/user_usage.csv")
user_device = pd.read_csv("data/user_device.csv")
android_devices = pd.read_csv("data/android_devices.csv")
```

```
In [147]: user_usage.head()
```

Out[147]:

	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id
0	21.97	4.82	1557.33	22787
1	1710.08	136.88	7267.55	22788
2	1710.08	136.88	7267.55	22789
3	94.46	35.17	519.12	22790
4	71.59	79.26	1557.33	22792

```
In [148]: user_device.head()
```

Out[148]:

	use_id	user_id	platform	platform_version	device	use_type_id
0	22782	26980	ios	10.2	iPhone7,2	2
1	22783	29628	android	6.0	Nexus 5	3
2	22784	28473	android	5.1	SM-G903F	1
3	22785	15200	ios	10.2	iPhone7,2	3
4	22786	28239	android	6.0	ONE E1003	1

```
In [149]: android_devices = android_devices.rename(index=str, columns={'Device': 'device'})
android_devices.head()
```

Out[149]:

	Retail Branding	Marketing Name	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

Merge using pandas

```
In [150]: usage_and_device = pd.merge(user_usage, user_device[['use_id', 'device']], on='use_id')
print('Total:', usage_and_device.shape[0])
usage_and_device.head()
```

Total: 159

Out[150]:

	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id	device
0	21.97	4.82	1557.33	22787	GT-I9505
1	1710.08	136.88	7267.55	22788	SM-G930F
2	1710.08	136.88	7267.55	22789	SM-G930F
3	94.46	35.17	519.12	22790	D2303
4	71.59	79.26	1557.33	22792	SM-G361F

```
In [151]: usage_and_device_android = pd.merge(usage_and_device,
        android_devices[['Model', 'Retail Branding']],
        left_on='device', right_on='Model').drop_duplicates()
print('Total:', usage_and_device_android.shape[0])
usage_and_device_android.head()
```

Total: 150

Out[151]:

	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id	device	Model	Retail Branding
0	21.97	4.82	1557.33	22787	GT-I9505	GT-I9505	Samsung
1	69.80	14.70	25955.55	22801	GT-I9505	GT-I9505	Samsung
2	249.26	253.22	1557.33	22875	GT-I9505	GT-I9505	Samsung
3	249.26	253.22	1557.33	22876	GT-I9505	GT-I9505	Samsung
4	83.46	114.06	3114.67	22880	GT-I9505	GT-I9505	Samsung

```
In [160]: usage_and_device_android.groupby('Retail Branding').agg({
    "outgoing_mins_per_month": "mean",
    "outgoing_sms_per_month": "mean",
    "monthly_mb": "mean",
    "use_id": "count"
}).sort_values('use_id', ascending=False)
```

Out[160]:

	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id
Retail Branding				
Samsung	196.975556	93.815354	3725.970707	99
HTC	289.315789	97.678421	7080.200000	19
Sony	143.703846	39.114615	2715.352308	13
Motorola	96.780000	68.844000	4195.424000	5
OnePlus	308.740000	51.772500	8824.890000	4
Huawei	81.526667	9.500000	1561.226667	3
LGE	111.530000	12.760000	1557.330000	2
Lava	60.650000	261.900000	12458.670000	2
Lenovo	215.920000	12.930000	1557.330000	1
Vodafone	42.750000	46.830000	5191.120000	1
ZTE	42.750000	46.830000	5191.120000	1

Merge using pandasql

```
In [178]: usage_and_device_sql = pds.sqldf("""
          SELECT uu.*, device FROM user_usage uu
          JOIN user_device USING (use_id)
          """, {'user_usage': user_usage, 'user_device': user_device})
print('Total:', usage_and_device_sql.shape[0])
usage_and_device_sql.head()
```

Total: 159

Out[178]:

	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id	device
0	21.97	4.82	1557.33	22787	GT-I9505
1	1710.08	136.88	7267.55	22788	SM-G930F
2	1710.08	136.88	7267.55	22789	SM-G930F
3	94.46	35.17	519.12	22790	D2303
4	71.59	79.26	1557.33	22792	SM-G361F

```
In [201]: usage_and_device_android_sql = pds.sqldf("""
          SELECT DISTINCT ud.*, `Retail Branding` FROM usage_and_device ud
          JOIN android_devices ad ON ud.device = ad.Model
          """, {'usage_and_device': usage_and_device_sql, 'android_devices': android_devices})
print('Total:', usage_and_device_android_sql.shape[0])
usage_and_device_android_sql.head()
```

Total: 150

Out[201]:

	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id	device	Retail Branding
0	21.97	4.82	1557.33	22787	GT-I9505	Samsung
1	1710.08	136.88	7267.55	22788	SM-G930F	Samsung
2	1710.08	136.88	7267.55	22789	SM-G930F	Samsung
3	94.46	35.17	519.12	22790	D2303	Sony
4	71.59	79.26	1557.33	22792	SM-G361F	Samsung

```
In [205]: pds.sqlldf("""
SELECT `Retail Branding`,
      AVG(outgoing_mins_per_month),
      AVG(outgoing_sms_per_month),
      AVG(monthly_mb),
      COUNT(use_id) use_id FROM usage_and_device_android
GROUP BY `Retail Branding`
ORDER BY use_id DESC
""", {'usage_and_device_android': usage_and_device_android_sql})
```

Out[205]:

	Retail Branding	AVG(outgoing_mins_per_month)	AVG(outgoing_sms_per_month)	AVG(monthly_mb)	use_id
0	Samsung	196.975556	93.815354	3725.970707	99
1	HTC	289.315789	97.678421	7080.200000	19
2	Sony	143.703846	39.114615	2715.352308	13
3	Motorola	96.780000	68.844000	4195.424000	5
4	OnePlus	308.740000	51.772500	8824.890000	4
5	Huawei	81.526667	9.500000	1561.226667	3
6	LGE	111.530000	12.760000	1557.330000	2
7	Lava	60.650000	261.900000	12458.670000	2
8	Lenovo	215.920000	12.930000	1557.330000	1
9	Vodafone	42.750000	46.830000	5191.120000	1
10	ZTE	42.750000	46.830000	5191.120000	1