In [29]:
```
import numpy as np
import pandas as pd
pd.set_option('display.max.columns', 100)
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 [30]:
```
user_usage = pd.read_csv("user_usage.csv")
user_device = pd.read_csv("user_device.csv")
android_devices = pd.read_csv("android_devices.csv")
```

In [31]:
```
user_usage.head()
```

Out[31]:

	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 [32]:
```
user_device.head()
```

Out[32]:

	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 [33]:
```
android_devices = android_devices.rename(index=str, columns={'Device': 'device'})
android_devices.head()
```

Out[33]:

	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 [34]:
```
user_usage_and_user_device = pd.merge(user_usage, user_device[['use_id', 'device']], on='use_id')
user_usage_and_user_device.head()
```

Out[34]:

	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 [35]:
```
user_usage_and_user_device_and_android_devices = pd.merge(user_usage_and_user_device,
                                    android_devices[['Model', 'Retail Branding']],
                                    left_on='device', right_on='Model').drop_duplicates()
user_usage_and_user_device_and_android_devices.head()
```

Out[35]:

	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 [36]:
```
user_usage_and_user_device_and_android_devices.groupby('Retail Branding').agg({
    "outgoing_mins_per_month": "mean",
    "outgoing_sms_per_month": "mean",
    "monthly_mb": "mean",
    "use_id": "count"
})
```

Out[36]:

Retail Branding	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb	use_id
HTC	289.315789	97.678421	7080.200000	19
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
Motorola	96.780000	68.844000	4195.424000	5
OnePlus	308.740000	51.772500	8824.890000	4
Samsung	196.975556	93.815354	3725.970707	99
Sony	143.703846	39.114615	2715.352308	13
Vodafone	42.750000	46.830000	5191.120000	1
ZTE	42.750000	46.830000	5191.120000	1

Merge using pandasql

In [37]:
```
query = """
    SELECT uu.*, ud.device FROM user_usage AS uu
    JOIN user_device AS ud ON uu.use_id = ud.use_id
"""
user_usage_and_user_device_sql = pds.sqldf(query, {'user_usage': user_usage, 'user_device': user_device})
user_usage_and_user_device_sql.head()
```

Out[37]:

	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 [38]:
```
query = """
    SELECT DISTINCT ud.*, ad.`Retail Branding` FROM user_usage_and_user_device AS ud
    JOIN android_devices AS ad ON ud.device = ad.Model
"""
user_usage_and_user_device_and_android_devices_sql = pds.sqldf(
    query, {'user_usage_and_user_device': user_usage_and_user_device_sql, 'android_devices': android_devices
})
user_usage_and_user_device_and_android_devices_sql.head()
```

Out[38]:

	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 [39]:
```
query = """
    SELECT `Retail Branding`,
           AVG(outgoing_mins_per_month) AS outgoing_mins_per_month,
           AVG(outgoing_sms_per_month) AS outgoing_sms_per_month,
           AVG(monthly_mb) AS monthly_mb,
           COUNT(use_id) AS use_id
           FROM user_usage_and_user_device_and_android_devices
    GROUP BY `Retail Branding`
"""
pds.sqldf(
    query, {'user_usage_and_user_device_and_android_devices': user_usage_and_user_device_and_android_devices_
sql})
```

Out[39]:

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