In [2]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import cufflinks as cf
plt.rcParams['font.sans-serif'] = ['Microsoft YaHei']
%matplotlib inline
```

一、读取数据

In [3]:

```
install =pd.read_excel(r"/Users/fennyw/Desktop/数据源整合.xlsx",sheet_name='安装信息')
registers = pd.read_excel(r"/Users/fennyw/Desktop/数据源整合.xlsx",sheet_name = "注册信
```

In [4]:

```
install.head()
```

Out[4]:

	安装时间	渠道	子渠道	地区	WIFI	用户唯一ID	系统	机型	操作系统版本
0	2020-04-26 15:59:27	渠道A	site01	地区A	True	1587916759000-8355351393884172615	android	samsung-SM-N900	5
1	2020-04-26 15:58:27	渠道A	site02	地区A	False	1587916702141-689636393710525296	android	samsung-SM-N960F	10
2	2020-04-26 15:56:57	渠道A	site01	地区A	False	1587916613722-2703192501000635621	android	samsung-SM-A7050	9
3	2020-04-26 15:50:55	渠道A	site03	地区A	True	1587916250955-4061104808165063458	android	OPPO-CPH1721	7
4	2020-04-26 15:49:42	渠道A	site04	地区A	True	1587916177009-3388800810186375808	android	HUAWEI-LYA-L29	9

In [5]:

```
registers.head()
```

Out[5]:

	用户类型	账号名称	安装时间	注册时间	注册渠道	子站	用户唯一ID	注册游戏	系统	机型
0	new	code03	2020-04-19 23:58:18	2020-04-20 00:06:51	渠道A	site12	1587311896496-9121742265690995358	游戏A	android	SM-G887F
1	old	code15	2020-04-20 00:04:51	2020-04-20 00:22:09	渠道A	site39	1587312290514-3372644182163270448	游戏A	android	SM-A730F
2	old	code17	2020-04-20 00:18:30	2020-04-20 00:25:18	渠道A	site32	1587313106069-4890529650028494525	游戏A	android	ASUS_I001DE
3	old	code19	2020-04-20 00:20:36	2020-04-20 00:27:39	渠道A	site40	1587313233938-6534515892410501030	游戏A	android	SM-G9880
4	old	code27	2020-04-20 00:28:29	2020-04-20 00:35:35	渠道A	site40	1587313708644-1610695342336039164	游戏A	android	SM-N9500

In [6]:

```
install.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7438 entries, 0 to 7437
Data columns (total 9 columns):
 #   Column   Non-Null Count  Dtype         
---  ------   --------------  -----         
 0   安装时间    7438 non-null   datetime64[ns]
 1   渠道      7438 non-null   object        
 2   子渠道     7438 non-null   object        
 3   地区      7438 non-null   object        
 4   WIFI    7438 non-null   bool          
 5   用户唯一ID  7438 non-null   object        
 6   系统      7438 non-null   object        
 7   机型      7438 non-null   object        
 8   操作系统版本  7438 non-null   int64         
dtypes: bool(1), datetime64[ns](1), int64(1), object(6)
memory usage: 472.3+ KB
```

In [7]:

```
install["安装时间"] = pd.to_datetime(install["安装时间"]).dt.date
install
```

Out[7]:

	安装时间	渠道	子渠道	地区	WIFI	用户唯一ID	系统	机型	操作系统版本
0	2020-04-26	渠道A	site01	地区A	True	1587916759000-8355351393884172615	android	samsung-SM-N900	5
1	2020-04-26	渠道A	site02	地区A	False	1587916702141-689636393710525296	android	samsung-SM-N960F	10
2	2020-04-26	渠道A	site01	地区A	False	1587916613722-2703192501000635621	android	samsung-SM-A7050	9
3	2020-04-26	渠道A	site03	地区A	True	1587916250955-4061104808165063458	android	OPPO-CPH1721	7
4	2020-04-26	渠道A	site04	地区A	True	1587916177009-3388800810186375808	android	HUAWEI-LYA-L29	9
...	...	...	...	...	...	...	...	...	...
7433	2020-04-19	渠道B	site255	地区B	False	1587313014578-7973769045124572018	android	samsung-SM-C9000	8
7434	2020-04-19	渠道B	site253	地区B	True	1587312902546-6172877259475361085	android	samsung-SM-A6050	8
7435	2020-04-19	渠道B	site249	地区A	True	1587312879321-95413188790982415	android	samsung-SM-N900U	5
7436	2020-04-19	渠道B	site249	地区A	True	1587312602910-5949698816134283004	android	xiaomi-Redmi Note 4	7
7437	2020-04-19	渠道B	site248	地区A	True	1587312593076-4528847716145296900	android	OPPO-CPH1917	9

7438 rows × 9 columns

In [8]:

```
registers.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5992 entries, 0 to 5991
Data columns (total 11 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   用户类型    5992 non-null   object
 1   账号名称    5992 non-null   object
 2   安装时间    5992 non-null   object
 3   注册时间    5992 non-null   object
 4   注册渠道    5992 non-null   object
 5   子站      5992 non-null   object
 6   用户唯一ID  5992 non-null   object
 7   注册游戏    5992 non-null   object
 8   系统      5992 non-null   object
 9   机型      5992 non-null   object
 10  地区      5992 non-null   object
dtypes: object(11)
memory usage: 515.1+ KB
```

In [9]:

```
registers["安装时间"] = pd.to_datetime(registers["安装时间"]).dt.date
registers["注册时间"] = pd.to_datetime(registers["注册时间"]).dt.date
```

In [10]:

```
registers.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5992 entries, 0 to 5991
Data columns (total 11 columns):
 #   Column  Non-Null Count  Dtype 
---  ------  --------------  ----- 
 0   用户类型    5992 non-null   object
 1   账号名称    5992 non-null   object
 2   安装时间    5992 non-null   object
 3   注册时间    5992 non-null   object
 4   注册渠道    5992 non-null   object
 5   子站      5992 non-null   object
 6   用户唯一ID  5992 non-null   object
 7   注册游戏    5992 non-null   object
 8   系统      5992 non-null   object
 9   机型      5992 non-null   object
 10  地区      5992 non-null   object
dtypes: object(11)
memory usage: 515.1+ KB
```

In [11]:

```
registers.head()
```

Out[11]:

	用户类型	账号名称	安装时间	注册时间	注册渠道	子站	用户唯一ID	注册游戏	系统	机型	地区
0	new	code03	2020-04-19	2020-04-20	渠道A	site12	1587311896496-9121742265690995358	游戏A	android	SM-G887F	地区A
1	old	code15	2020-04-20	2020-04-20	渠道A	site39	1587312290514-3372644182163270448	游戏A	android	SM-A730F	地区A
2	old	code17	2020-04-20	2020-04-20	渠道A	site32	1587313106069-4890529650028494525	游戏A	android	ASUS_I001DE	地区A
3	old	code19	2020-04-20	2020-04-20	渠道A	site40	1587313233938-6534515892410501030	游戏A	android	SM-G9880	地区B
4	old	code27	2020-04-20	2020-04-20	渠道A	site40	1587313708644-1610695342336039164	游戏A	android	SM-N9500	地区B

In [12]:

```
registers["激活时间"] = registers["注册时间"] - registers["安装时间"]
registers.head()
```

Out[12]:

	用户类型	账号名称	安装时间	注册时间	注册渠道	子站	用户唯一ID	注册游戏	系统	机型	地区	
0	new	code03	2020-04-19	2020-04-20	渠道A	site12	1587311896496-9121742265690995358	游戏A	android	SM-G887F	地区A	(
1	old	code15	2020-04-20	2020-04-20	渠道A	site39	1587312290514-3372644182163270448	游戏A	android	SM-A730F	地区A	(
2	old	code17	2020-04-20	2020-04-20	渠道A	site32	1587313106069-4890529650028494525	游戏A	android	ASUS_I001DE	地区A	(
3	old	code19	2020-04-20	2020-04-20	渠道A	site40	1587313233938-6534515892410501030	游戏A	android	SM-G9880	地区B	(
4	old	code27	2020-04-20	2020-04-20	渠道A	site40	1587313708644-1610695342336039164	游戏A	android	SM-N9500	地区B	(

三、游戏数据指标拆解与分析

1.激活

In [13]:

```
#提取激活时间列中的"days"前的数值
registers["激活时间"] = registers["激活时间"].astype(str).str.split("days").str[0]
```

In [14]:

```
registers["激活时间"] = registers["激活时间"].astype(int)
```

In [16]:

```
table = registers["激活时间"].groupby(pd.cut(registers["激活时间"],bins=[-1,0,1,3,7,10(
table
```

Out[16]:

```
激活时间
(-1, 0]      4877
(0, 1]        307
(1, 3]        128
(3, 7]        106
(7, 100]      574
Name: 激活时间, dtype: int64
```

In [18]:

```
table.reset_index(name="count")
```

Out[18]:

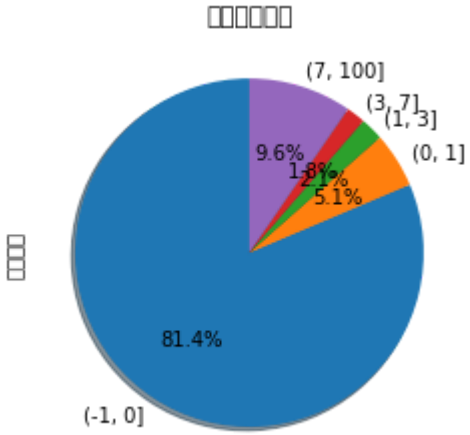
	激活时间	count
0	(-1, 0]	4877
1	(0, 1]	307
2	(1, 3]	128
3	(3, 7]	106
4	(7, 100]	574

In [19]:

```
table.plot(kind="pie",autopct="%1.1f%%",shadow=True,startangle=90,title="激活时间占比"
plt.show()
```

```
findfont: Font family ['sans-serif'] not found. Falling back to DejaVu
Sans.
/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 28608 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 27963 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 26102 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 38388 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 21344 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 27604 missing from current font.

findfont: Font family ['sans-serif'] not found. Falling back to DejaVu
Sans.
/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 28608 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 27963 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 26102 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 38388 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:
```

```
Glyph 21344 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 27604 missing from current font.
```



2.渠道质量

渠道的获客数量、获客质量以及获客成本是我们评价渠道质量好坏的主要参数。一个好的渠道应具备流量较为充足、获客稳定、获客成本较低、各个环节的转化率较好这几个条件。

In [20]:

```
#渠道的每日获客数量趋势
时间_渠道_count = pd.pivot_table(install,index=["安装时间"],values=["用户唯一ID"],columr
时间_渠道_count
```

Out[20]:

	用户唯一ID	
渠道	渠道A	渠道B
安装时间		
2020-04-19	41	115
2020-04-20	230	717
2020-04-21	225	687
2020-04-22	252	793
2020-04-23	236	786
2020-04-24	203	712
2020-04-25	340	938
2020-04-26	316	847

In [21]:

```
时间_渠道_count.iplot(kind="line",title="各渠道的每日获客量")
```

```
e() * 1000)) - start_time

~/opt/anaconda3/lib/python3.7/site-packages/retrying.py in get(self,
 wrap_exception)
    245                     raise RetryError(self)
    246                 else:
--> 247                     six.reraise(self.value[0], self.value[1], sel
f.value[2])
    248             else:
    249                 return self.value

~/opt/anaconda3/lib/python3.7/site-packages/six.py in reraise(tp, val
ue, tb)
    701                 if value.__traceback__ is not tb:
    702                     raise value.with_traceback(tb)
--> 703                 raise value
    704             finally:
    705                 value = None

~/opt/anaconda3/lib/python3.7/site-packages/retrying.py in call(self,
```

In [27]:

```
道=='渠道A']['激活时间'].groupby(pd.cut(registers[registers.注册渠道=='渠道A']['激活时间']
```

Out[27]:

```
激活时间
(-1, 0]     1356
(0, 1]        69
(1, 3]        24
(3, 7]        18
(7, 100]      28
Name: 激活时间, dtype: int64
```

In [36]:

```
table_B=registers[registers.注册渠道=='渠道B']['激活时间'].groupby(pd.cut(registers[reg
table_B
```

Out[36]:

```
激活时间
(-1, 0]     3521
(0, 1]       238
(1, 3]       104
(3, 7]        88
(7, 100]     546
Name: 激活时间, dtype: int64
```

In [38]:

```
table_A.reset_index(name="count_A")
```

Out[38]:

	激活时间	count_A
0	(-1, 0]	1356
1	(0, 1]	69
2	(1, 3]	24
3	(3, 7]	18
4	(7, 100]	28

In [39]:

```
table_B.reset_index(name="count_B")
```

Out[39]:

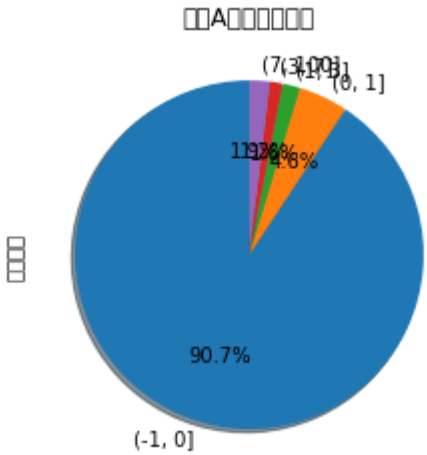
	激活时间	count_B
0	(-1, 0]	3521
1	(0, 1]	238
2	(1, 3]	104
3	(3, 7]	88
4	(7, 100]	546

In [40]:

```
table_A.plot(kind='pie',autopct='%1.1f%%',shadow=True,startangle=90,title='渠道A激活时
plt.show()
```

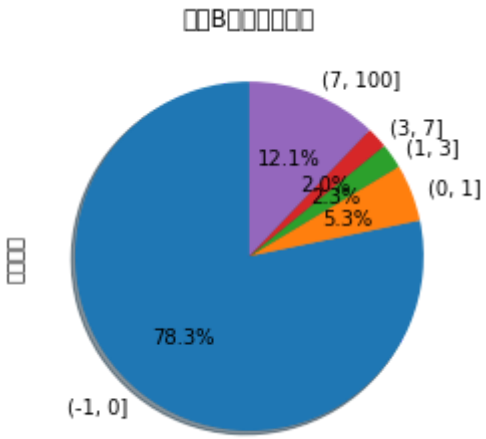
```
/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 28192 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 36947 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 28608 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 27963 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 26102 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 38388 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 21344 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:211: RuntimeWarning:

Glyph 27604 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 28608 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 27963 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 26102 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
```

```
kends/backend_agg.py:180: RuntimeWarning:

Glyph 38388 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 28192 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 36947 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 21344 missing from current font.

/Users/fennyw/opt/anaconda3/lib/python3.7/site-packages/matplotlib/bac
kends/backend_agg.py:180: RuntimeWarning:

Glyph 27604 missing from current font.
```



In [41]:

```
table_B.plot(kind='pie',autopct='%1.1f%%',shadow=True,startangle=90,title='渠道B激活时
plt.show()
```



渠道A激活时间大于7天的比例为1.92%，渠道B激活时间大于7天的比例为12.1%。渠道B当天激活的用户比例也低于渠道A。可见，渠道B的获客质量要比渠道A的获客质量差一些。

综上所述，渠道A的获客数量<渠道B，渠道A的获客质量>渠道B。市场推广部门可以适当加大对于渠道A的广告投入，多对渠道A做一些获客测试，提升渠道A的获客数量

3.用户画像

In [43]:

```
#提取安装信息表中"机型"列的文字
install["机型"] = install["机型"].str.split("-").str[0]
install["机型"].replace(["Xiaomi","Galaxy"],["xiaomi","samsung"],inplace=True)
phone_model = pd.pivot_table(install,index="机型",values="用户唯一ID",aggfunc=len)
phone_model.sort_values(by="用户唯一ID",inplace=True,ascending=True)
phone_model
```

Out[43]:

机型	用户唯一ID
8848	1
RYJJ	1
RUIO	1
Qua	1
QMX	1
...	...
asus	587
xiaomi	624
HUAWEI	726
OPPO	1150
samsung	2703

139 rows × 1 columns

In [45]:

```
region_table = pd.pivot_table(install,index="地区",values="用户唯一ID",aggfunc=len)
region_table
```

Out[45]:

地区	用户唯一ID
地区A	5764
地区B	1596
地区C	78

从用户地域来看，地区A的用户数量最多，地区B的用户数量其次，地区C的用户数量很少。

游戏的推广及广告投放中，对地域进行定向，避免在地区C耗费过多的广告投入。至于用户机型方面，安装用户数量排名前四的机型与目前安卓市场份额占比一致，不必对用户机型作广告定向设置。

四、总结

1.激活时间超过7天的用户占比较高，有被刷量的风险，这类用户注册率的改善需要进一步的数据支撑以及与业务部门的沟通；

2.渠道A的流量质量要好于渠道B，后续的推广中资源可以适当向渠道A做一些倾斜；

3.游戏广告投放中对地区A和地区B进行区域定向，减少预算在地区C上的浪费；

4.由于缺少游戏APP的其它用户数据，这里就只做市场推广相关指标模块的简要分析，用户活跃&留存指标、用户付费相关指标这里不进行分析。

In [ ]:

```

```