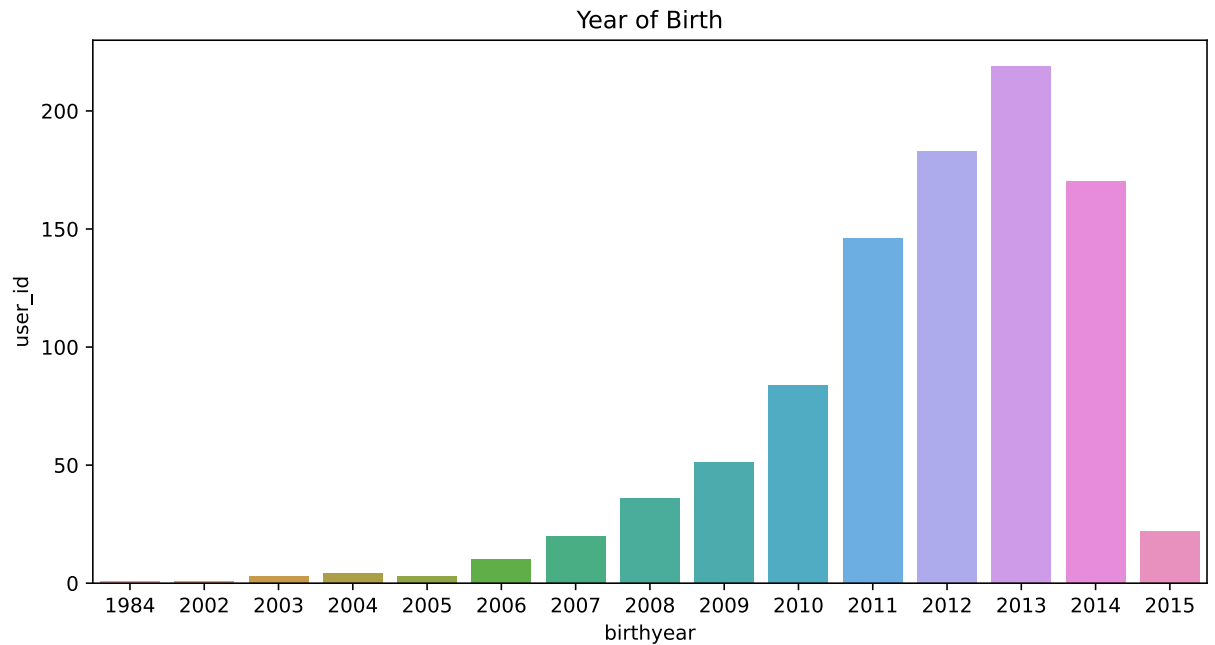
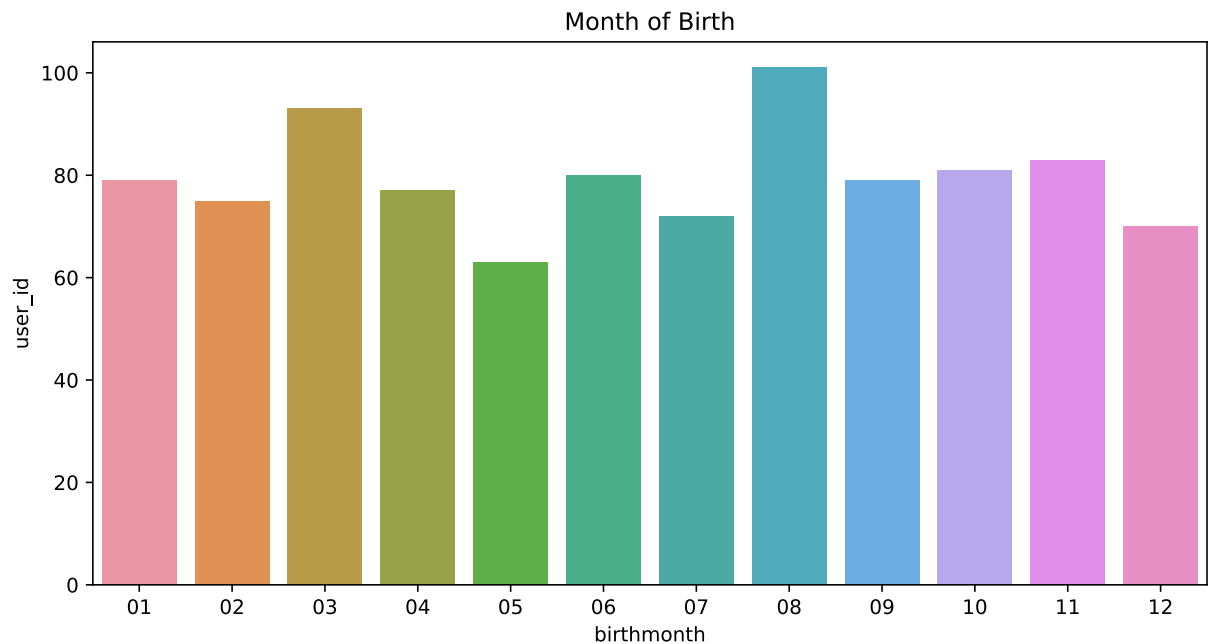

1. 分析用户特征

1.1 婴幼儿生日时间分布

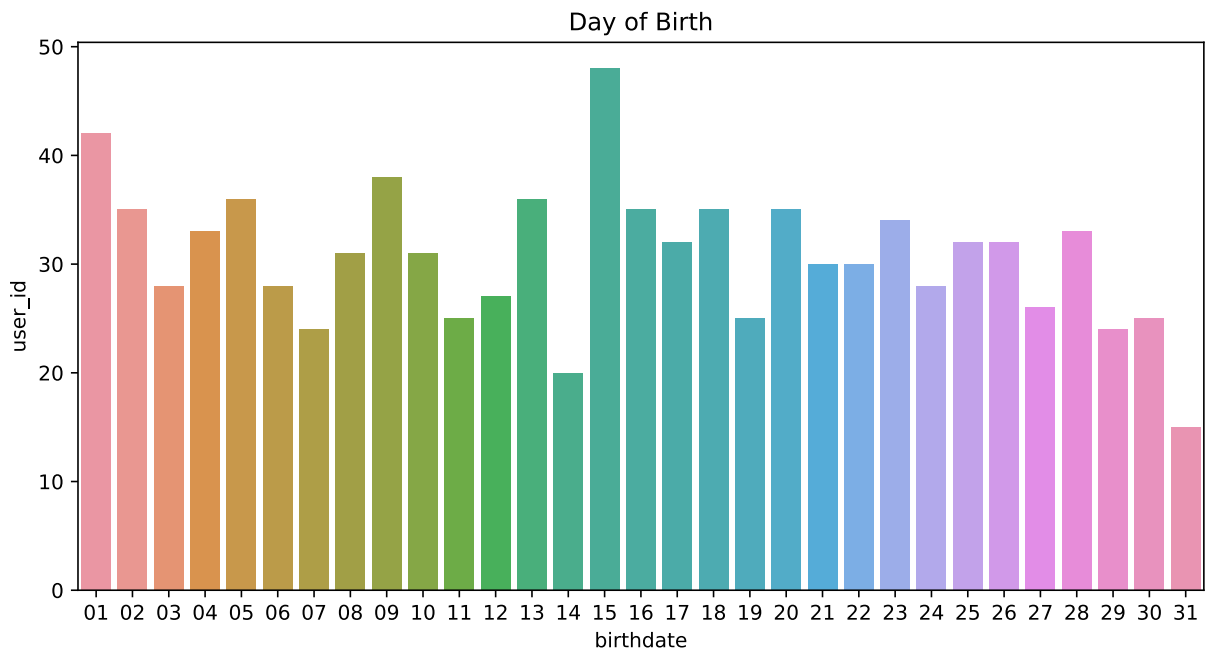
```
Out[ ]: <AxesSubplot:title={'center':'Year of Birth'}, xlabel='birthyear', ylabel='user_id'>
```



```
Out[ ]: <AxesSubplot:title={'center':'Month of Birth'}, xlabel='birthmonth', ylabel='user_id'>
```



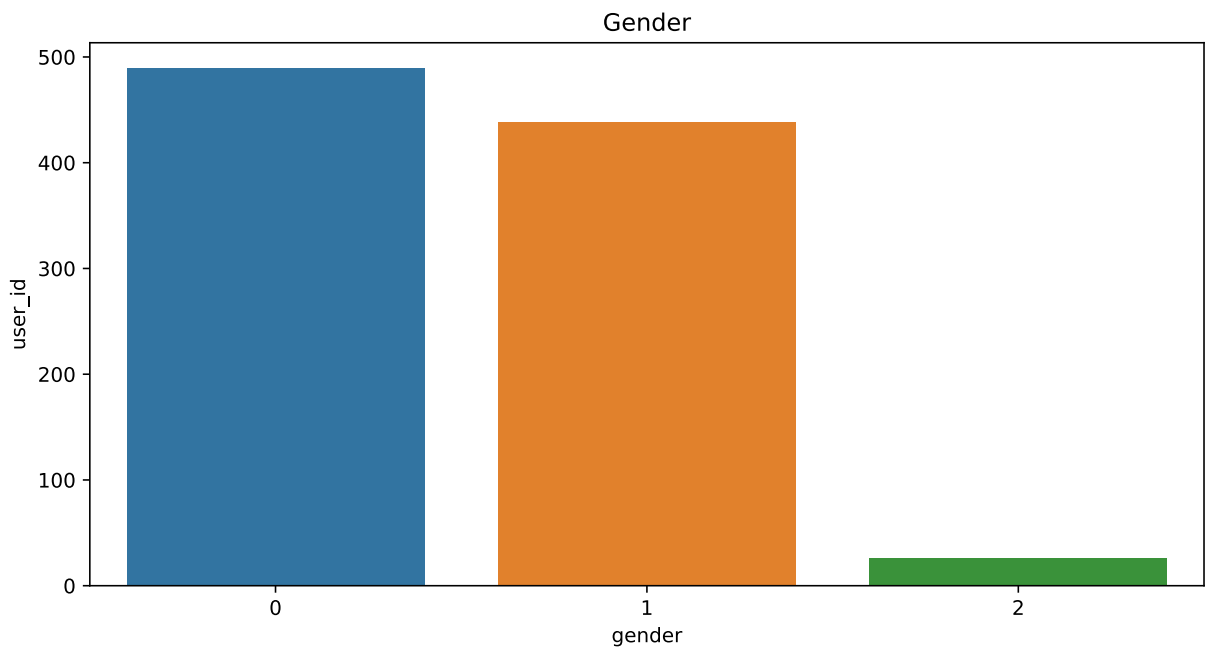
```
Out[ ]: <AxesSubplot:title={'center':'Day of Birth'}, xlabel='birthdate', ylabel='user_id'>
```



可以看出，对于为孩子购买商品的父母，他们孩子的出生时间主要集中在2012~2014年，而出生的月份和日期比较平均。具体的孩子年龄分布见后续分析。

1.2 婴幼儿性别分布

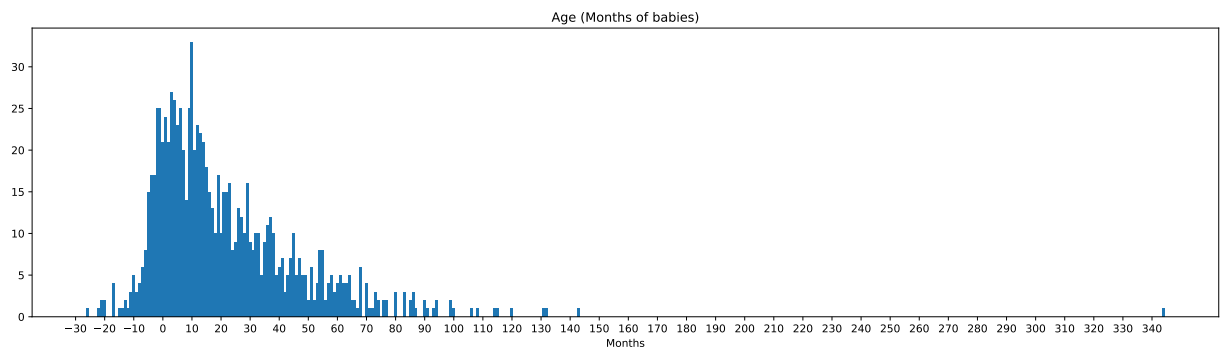
```
Out[ ]: <AxesSubplot:title={'center':'Gender'}, xlabel='gender', ylabel='user_id'>
```



显然，顾客中性别记为0的婴幼儿明显多于性别标记为1的，这可能是总体婴幼儿性别分布的体现。

1.2 婴幼儿年龄分布

```
Out[ ]: Text(0.5, 0, 'Months')
```

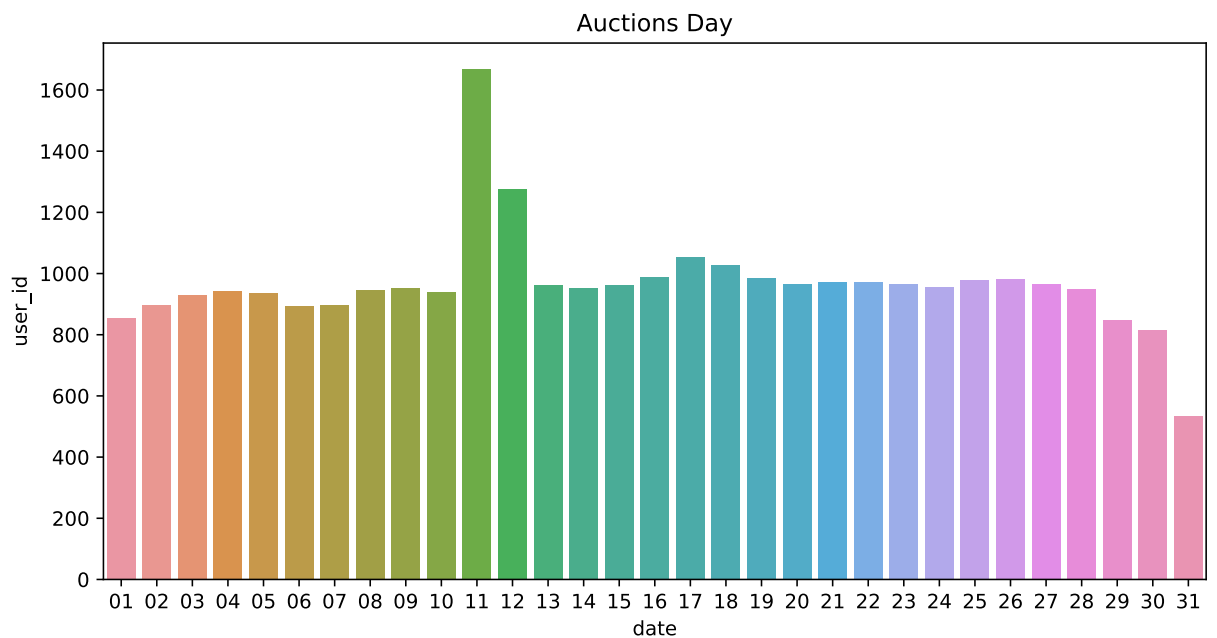


可以从图中看出，购买商品的父母，他们购买商品时一般是在孩子出生前一年到出生后两年中。可以以此为标准估计顾客消费的可能性和重要性。

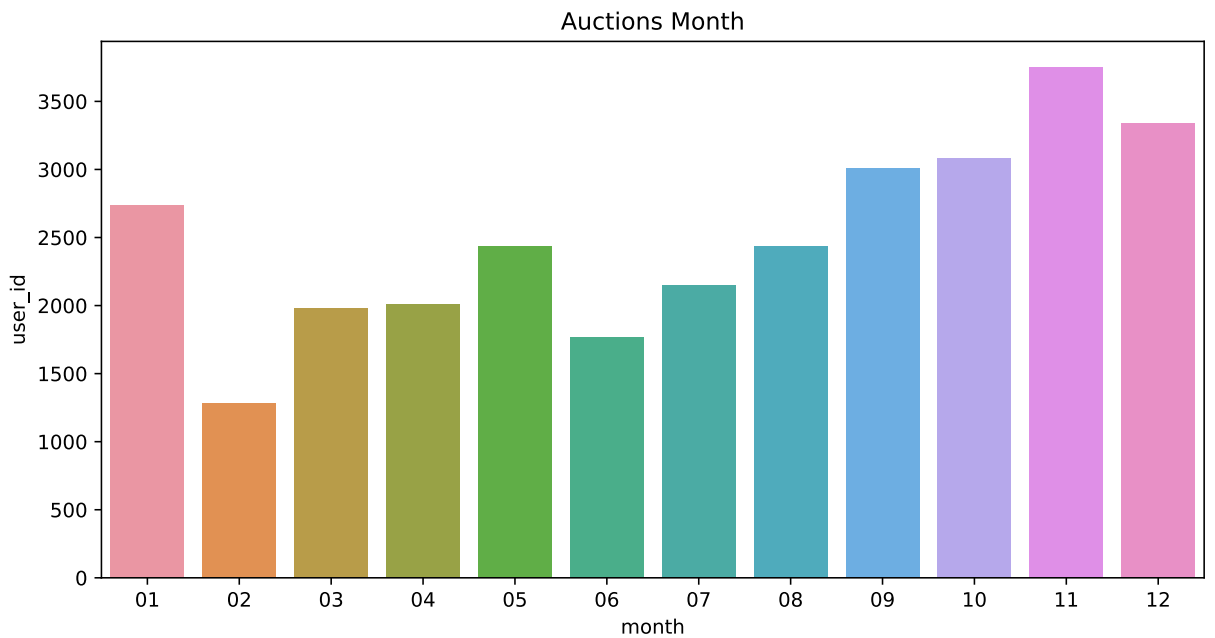
2. 分析订单特征

2.1 时间分布特征

```
Out[ ]: <AxesSubplot:title={'center':'Auctions Day'}, xlabel='date', ylabel='user_id'>
```



```
Out[ ]: <AxesSubplot:title={'center':'Auctions Month'}, xlabel='month', ylabel='user_id'>
```



2.2 客户重要性分析

根据销量确定的最重要20位客户

```
user_id
2288344467    10000
117730165     2800
173701616     2748
1945590674     1500
32141414      1000
119395773      700
300857121      600
300205516      600
462029374      500
1671630112     498
105969610      450
125495768      410
119491758      400
50880819       399
871858982      350
1681976532     340
51409972       306
1925405730     300
10339114       300
65695459       276
Name: buy_mount, dtype: int64
```

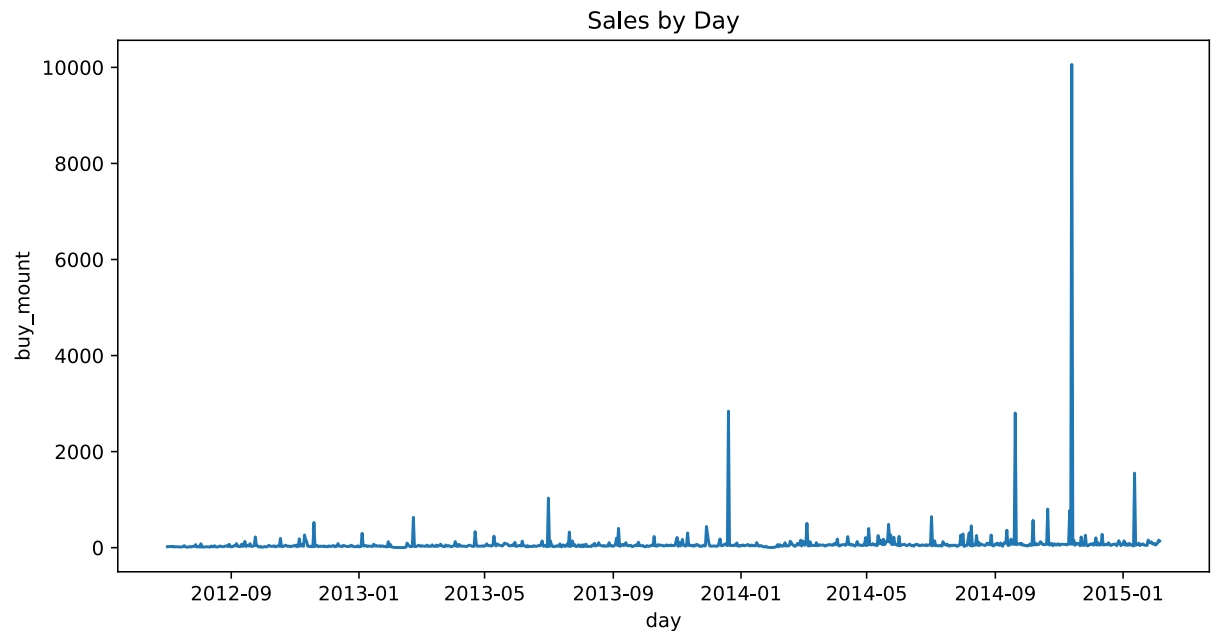
根据交易次数确定的最重要20位客户

```
user_id
814316568     4
69889555      2
718313786     2
213455117     2
21833701      2
890739828     2
371804010     2
473068583     2
116466705     2
900655351     2
1954450396    2
14069200      2
613315428     2
775005980     2
1126907662    2
```

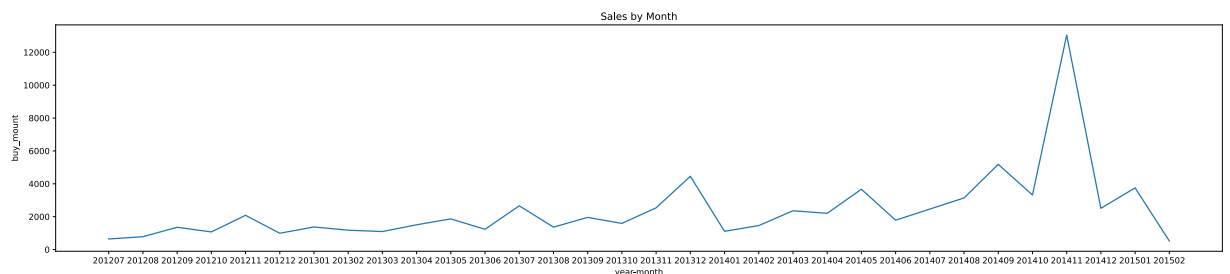
```
889177953      2
1137719147      2
50570082        2
57700536        2
73181079        2
Name: buy_mount, dtype: int64
```

2.3 销量统计

```
Out[ ]: <AxesSubplot:title={'center':'Sales by Day'}, xlabel='day', ylabel='buy_mount'>
```



```
Out[ ]: <AxesSubplot:title={'center':'Sales by Month'}, xlabel='year-month', ylabel='buy_mount'>
```

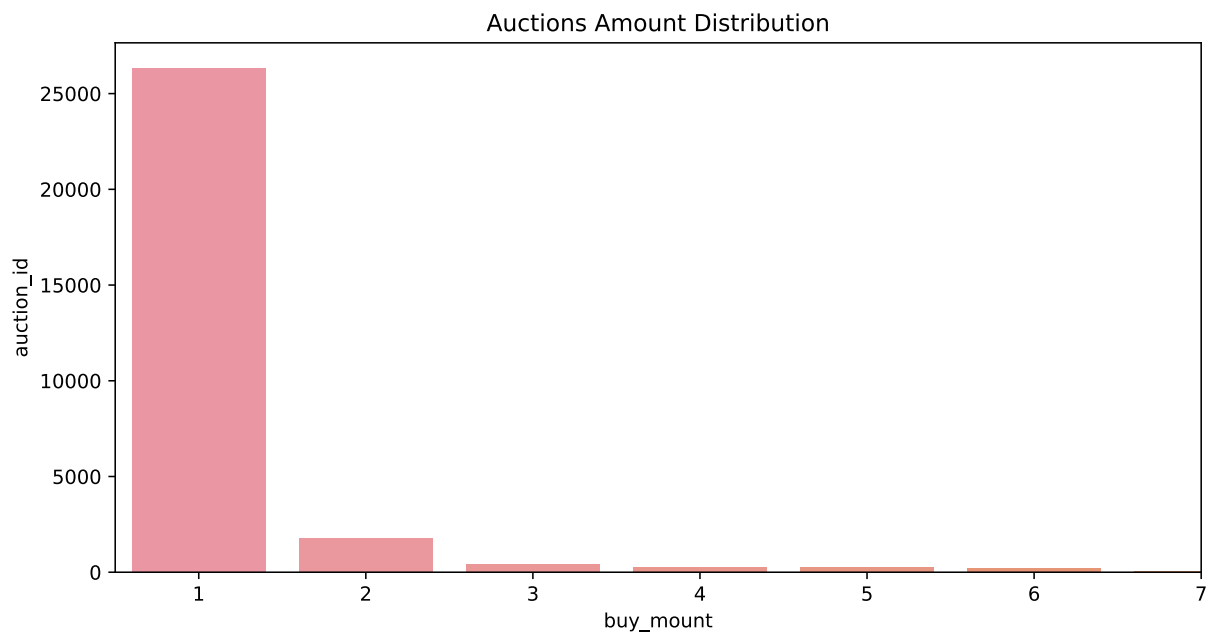


可以看出单日销量总体平稳，更明显的增长体现在双十一等购物节的增长上。

3. 交易和商品特征

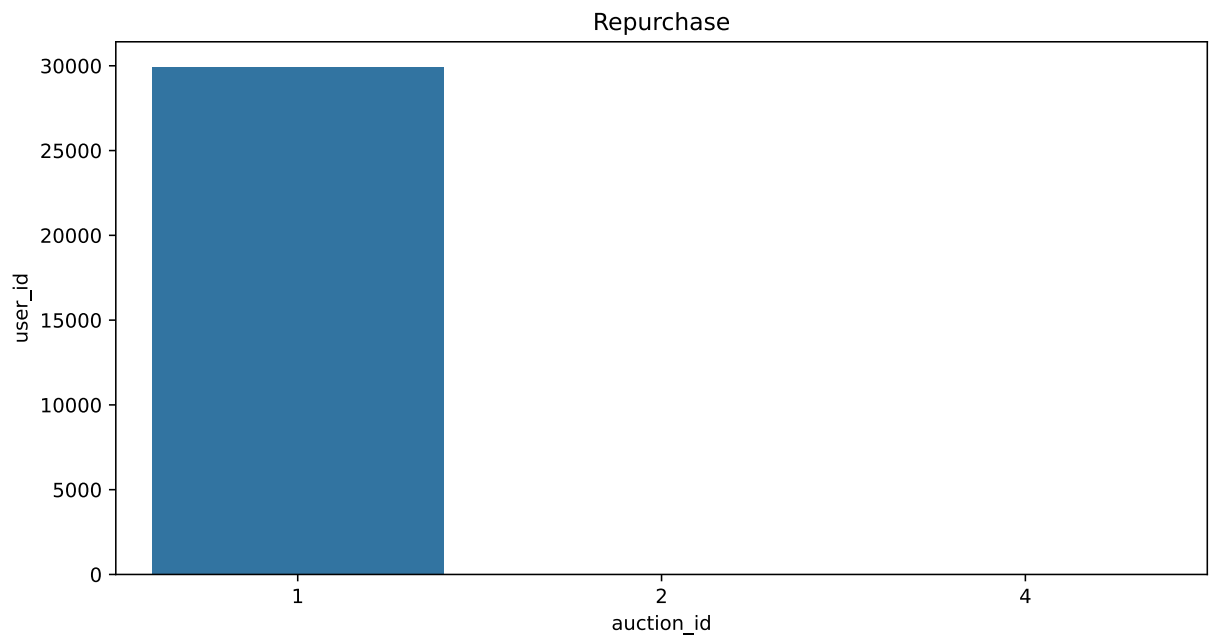
3.1 单次购买数量分布

```
Out[ ]: (-0.5, 6.0)
```



可以看出主要是零售业务

Out[]: <AxesSubplot:title={'center': 'Repurchase'}, xlabel='auction_id', ylabel='user_id'>



同时也几乎没人复购

3.2 商品品类热度排名

交易品类 (cat_id) 排名

cat_id	
50013636	2066
50010558	1341
50013207	1148
50006602	924
50011993	864
50012451	739
50018831	718
50012359	554
50010555	522
211122	502
50023591	476
50010566	464

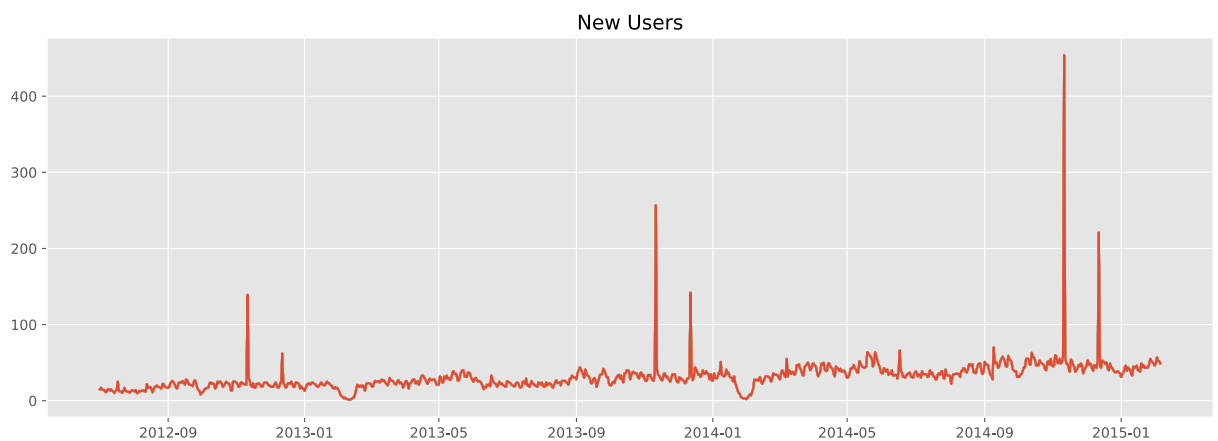
```
50013711    431
50010549    420
50010557    413
50003700    407
50012365    390
50007011    361
50010544    331
50012442    321
Name: auction_id, dtype: int64
```

交易品类 (cat1) 排名

```
cat1
50008168    12494
28          6963
50014815     4834
50022520     2367
122650008     2110
38          1203
Name: auction_id, dtype: int64
```

3.1 新用户变化

```
D:\programs\envs\python38\lib\site-packages\pandas\core\apply.py:577: FutureWarning: Dropping invalid columns in DataFrameGroupBy.min is deprecated. In a future version, a
TypeError will be raised. Before calling .min, select only columns which should be valid for the function.
    return f(*args, **kwargs)
```



可以看出总体上新用户数量缓慢增长，而在购物节等促销活动中新用户出现爆发性增长。