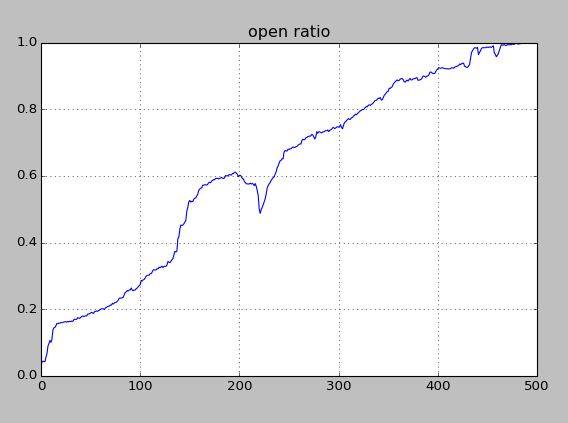
2017-02-01

本文主要基于count\_user\_pay.csv进行分析

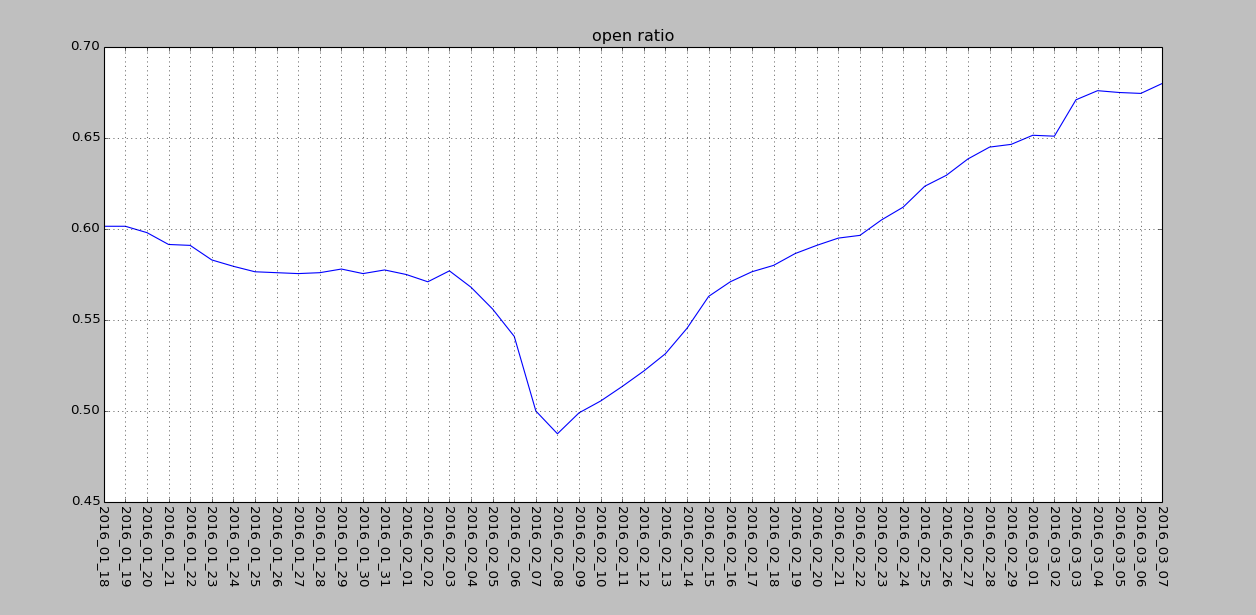
>>> open\_ratio():

用于分析商家在第[0,488]天（对应2015-07-01 ~ 2016-10-31）中的开张率。例如：图（100,0.23）的含义是在第100天，2000家商铺大概有2000\*0.23家开张。（ “开张率”是形象的称呼，实际上应当是 2000\*0.23的商家有客流，其余的在这一天都没有客流。）



观察上图可以发现，在第200~250天之间，开张率出现明显的下滑，我将观察范围缩小。

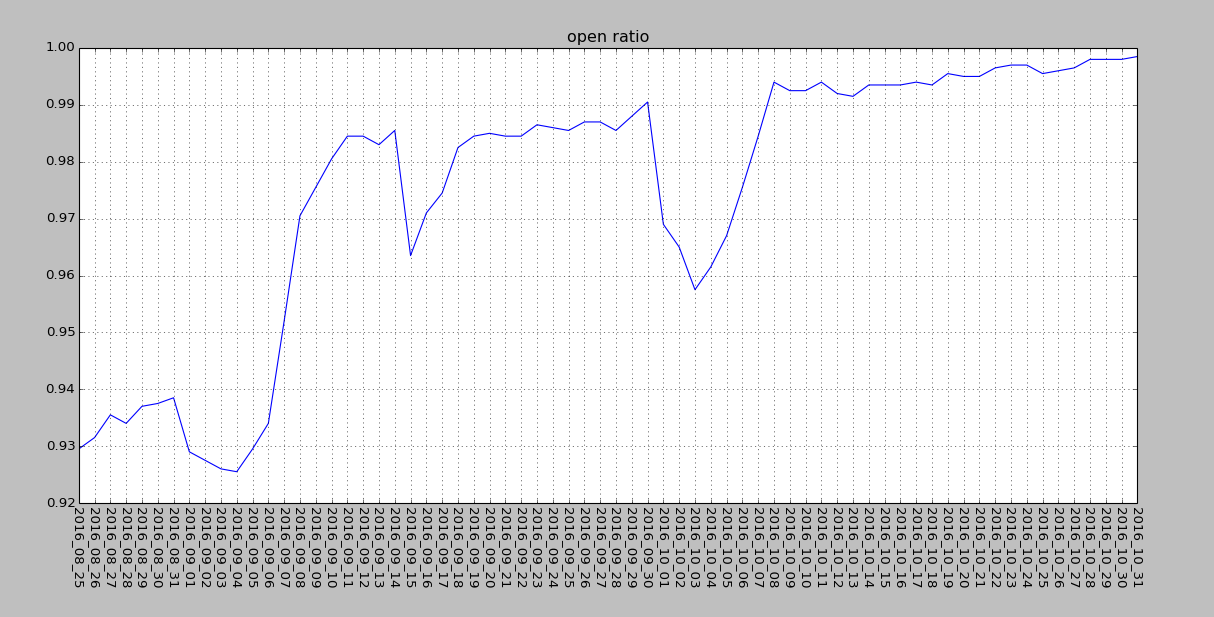
>>> open\_ratio(200,250)



观察范围缩小之后，发现波谷是2016-02-08（春节）。因此，在春节前后，一部分商家选择关门或者客流量为0。（为了显示清晰，只有当查找范围小于100天，既 250-200<100 的时候，横坐标才会显示日期。）

在第420天到最尾，开张率几乎稳定在峰值。放大观察：

>>> open\_ratio(420)



从2016-08-31 ~ 2016-09-11 经历了先下滑，后上涨（我个人觉得可能跟学生的开学有关）。

2016-09-15（中秋节）到达波谷。

2016-10-01 ~ 2016-10-07（国庆节）也对开张率产生较大影响。

从2016-10-08开始，开张率稳定在99%以上，既每天还有<200家的商家客流量为0.

综上：重大的国家法定节假日会对开张率产生较大影响，但是由于待预测范围：2016-11-01 ~ 2016-11-14不存在法定节假日，因此需要注意的是：取训练集的时候应当尽量避开这些日期范围。