**任务：**对CDNow网站的用户购买明细数据进行消费行为分析。

**一、加载数据**

1、创建表头，加载txt数据：read\_csv()

2、数据初步统计：describe()

3、转换时间类型：to\_datetime()、astype(‘datetime64[M]’)

**二、多维度数据统计**

1、从用户角度分析：groupby()、sum()

2、按月销量、金额统计：groupby()、sum()

数据异常，分析原因：

（1）分析销量和金额关系

a 绘制销量、金额散点图：scatter()

b 以用户分组绘制销量、金额散点图：groupby()、scatter()

c 绘制金额、销量直方图：hist()

（2）分析消费时间节点

a 第一次消费时间：min()、value\_counts()

b 最后一次消费时间：max()、value\_counts()

总结：大部分用户消费集中在前三月，导致数据早期销售额多的原因。

**3、复购率（同一时间窗口购买两次以上）**

（1）计算用户每月购买量并数据透视：pivot\_table()

（2）数据转换：applymap(lambda x:1 if x>1 else np.NaN if x ==0 else 0)

（3）复购率：sum()/count()

**4、回购率（某个时间窗口有消费，下个时间窗口仍然消费）**

（1）计算用户每月购买平均金额并数据透视：pivot\_table()

（2）数据转换：applymap(lambda x:1 if x>0 else 0)、apply(purchase\_return)

purchase\_return为自建函数进行横向判断

（3）回购率：sum()/count()

**5、用户分层分析**

（1）创建函数判断用户层次：以本月是否有消费为分界

（2）排除“未来新客”，统计不同层次用户：value\_counts()

（3）计算回流用户和活跃用户占比：x/x.sum()

**6、用户质量分析**

（1）消费金额升序，累加并计算比例：sort\_values()、cumsum()

（2）销量升序，累加并计算比例：sort\_values()、cumsum()

**7、用户生命周期分析 注：数值换算**

（1）计算首次和最后一次消费时间：min()、max()

（2）计算用户平均生命周期：max()-min()、mean()

（3）排除一次消费用户，计算两次以上消费用户生命周期

**8、留存率（二次消费比率，多个时间窗口，倾向于计算第一次消费）**

（1）计算每一次消费距第一次消费时间差

（2）日期转换：timedelta64()

（3）数据分桶：cut()

（4）数据转换：lambda x:1 if x>0 else 0

（5）计算留存率：sum()/count()

（6）计算用户平均购买周期