

原

Python药店销售数据分析

2018年09月04日 01:38:42 huangxiaoyun1900 阅读数 1932

分析目标：根据药店销售数据，分析药品销售关键指标，以及药品销售趋势

1. 导入并清理数据

1import pandas as pd

2

3# 以object形式输入数据可保持数据原始形状，之后可用astype() 转换数据格式

4sale_data=pd.read_excel('药店2018年销售数据.xlsx',0,dtypes='object')

登录后复制

<

>

	购药时间	社保卡号	商品编码	商品名称	销售数量	应收金额	实收金额
0	2018-01-01 星期五	001616528	236701	强力VC银翘片	6	82.8	69
1	2018-01-02 星期六	001616528	236701	清熱解毒口服液	1	28	24.64
2	2018-01-06 星期三	0012602828	236701	感康	2	16.8	15
3	2018-01-11 星期一	0010070343428	236701	三九感冒灵	1	28	28
4	2018-01-15 星期五	00101554328	236701	三九感冒灵	8	224	208

1# 1. 列名重置：更改购药时间为销售时间

2dfrename = {"购药时间":"销售时间"}

3sale_data.rename(columns=dfrename,inplace=True)

4

5# 2. 处理缺失值

6# 查找缺失值

7sale_data.isnull().sum()

8

9# 删除缺失值

10sale_df=sale_data.dropna(subset=['销售时间','社保卡号'],how='any',axis=0)

11sale_df.shape

12

13# 删除缺失值后需重设索引，加drop=True可将原索引删除，否则原索引将保存为列名为index的一列

14sale_df=sale_df.reset_index(drop=True)

15

16# 3. 提取数据并转换数据格式

17# 将销售数量，应收金额，实收金额转换成浮点型

18sale_df['销售数量']=sale_df['销售数量'].astype('float')

19sale_df['应收金额']=sale_df['应收金额'].astype('float')

20sale_df['实收金额']=sale_df['实收金额'].astype('float')

21sale_df.dtypes

22

23# 4. 处理日期数据

24# 提取销售时间

25date = pd.DataFrame(sale_df['销售时间'].apply(lambda s: s.split(" ")).values.tolist(),columns=['日期','星期'])

26# 将日期的object类型转换成日期类型

27# 加errors='coerce',如果原始数据不符合日期的格式，转换后的值为空值NaT

28date['日期']=pd.to_datetime(date['日期'],format='%Y-%m-%d',errors='coerce')

29sale_df1 = sale_df

30sale_df1['销售日期']=date['日期']

31sale_df1['销售星期']=date['星期']

32

33# 转换日期格式时，不符合日期格式的会转换成空值，删除空值行

34sale_df1 = sale_df1.dropna(subset=['销售日期'],how='any')

35# 重置index

36sale_df1=sale_df1.reset_index(drop=True)

37

38# 5. 按日期排序

39sale_df1=sale_df1.sort_values(by='销售日期',ascending=True)

40# 重置index

41sale_df1=sale_df1.reset_index(drop=True)

42

43# 6. 异常值处理

44# 查看有无异常值

VIP

QR

```
45 sale_df1.describe()46 # “销售数量”，“金额” 最小值不能低于0
47 print("删除异常值前:",sale_df1.shape)
48 querySer = sale_df1.loc[:, '销售数量']>0
49 sale_df2 = sale_df1.loc[querySer,: ]
50 print("删除异常值后:",sale_df2.shape)
```

2. 分析数据

a. 计算关键指标：每月平均消费次数，每月平均消费金额，客单价

```
1 '''
2 业务指标1：月均消费次数=每月消费次数平均值
3 条件：同一人同一天发生的所有消费视作一次
4 方法：“销售日期”和“社保卡号”若同时相同，则只保留一项计数
5 '''
6 kpi1 = sale_df2.drop_duplicates(subset=[ '销售日期', '社保卡号'])
7 # 按时间顺序排序
8 kpi1 = kpi1.sort_values(by='销售日期',ascending=True)
9 # 重置索引
10 kpi1=kpi1.reset_index(drop=True)
11 kpi1.shape
12
13 kpi1['消费次数']=1
14 kpi = kpi1
15 kpi.index=kpi[ '销售日期' ]
16
17 # 用groupby统计每月消费数据
18 gb=kpi.groupby(kpi.index.month)
19 MonSales=gb.sum()
20
21
22 # 月均消费次数 kpi_1
23 kpi_1 = MonSales['消费次数'].loc[:6].mean()
24
25 # 月均销售金额 kpi_2
26 kpi_2 = MonSales['实收金额'].loc[:6].mean()
27
28 # 客单价 kpi_3
29 kpi_3 = MonSales['实收金额'].sum()/MonSales['消费次数'].sum()
```

	销售数量	应收金额	实收金额	消费次数
销售日期				
1	1966.0	41167.0	37846.22	845
2	1453.0	31035.2	28560.18	602
3	1832.0	35968.1	32851.77	823
4	2431.0	43433.2	38746.41	1044
5	1820.0	41777.2	38017.91	771
6	1889.0	42452.1	39107.26	750
7	1224.0	27227.9	25194.97	510

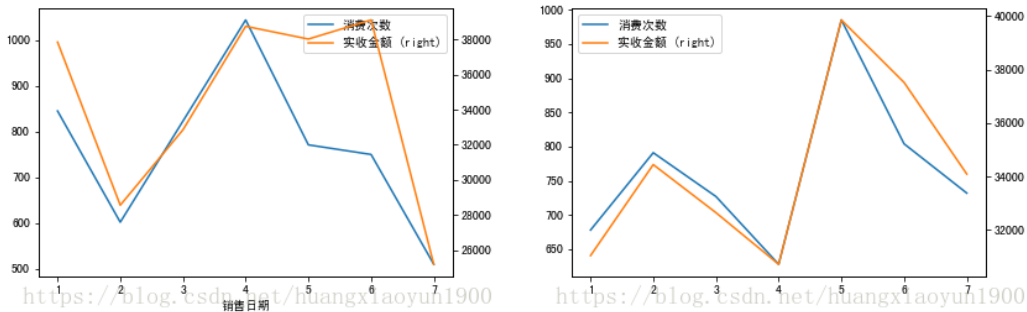
b. 比较各时段消费情况：

```
1 import matplotlib.pyplot as plt
2 %matplotlib inline
3 plt.rcParams['font.sans-serif']=['SimHei']
4
5 # 每月消费情况对比
6 f,ax1 = plt.subplots()
7 MonSales.plot(x=MonSales.index,y=[ '消费次数'],ax=ax1)
8 MonSales.plot(x=MonSales.index,y=[ '实收金额'],ax=ax1,secondary_y=True)
9
10 # 星期数据处理
11 kpi2 = kpi1
12 gb2 = kpi2.groupby(by='销售星期')
```

```

13 WeekSales=gb2.sum() 14 WeekSales.index=[1,3,2,5,6,4,7]
15 WeekSales.sort_index(inplace = True)
16
17 # 星期消费情况对比
18 f,ax2 = plt.subplots()
19 WeekSales.plot(x=WeekSales.index,y='消费次数',ax=ax2)
20 WeekSales.plot(x=WeekSales.index,y='实收金额',ax=ax2,secondary_y=True)

```



由上图可知:

1. 在月份销售对比中, 2月药店销售金额较其他月有明显的下降 (7月数据只到7月19日, 故不做整体比较), 4, 5, 6月药店销售金额均在38000以上。
2. 在月份销售对比中, 4月消费次数最多, 但实收金额并不是最高, 说明客单次消费相比6月较低。
3. 在星期销售对比中, 消费次数和消费金额总体趋势一致, 周五为全星期消费次数和消费金额之最, 周六次之。周四则消费最少。

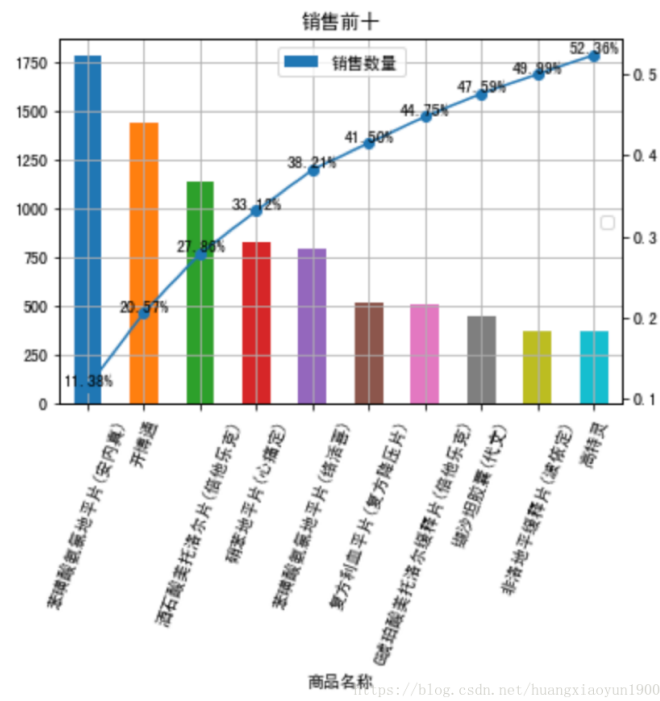
c. 销售药品统计

帕累托图

```

1 # 根据销售商品名称groupby, 排序销售数量, 选出Top10
2 gb3 = pd.DataFrame(sale_df2.groupby(['商品名称']).sum()['销售数量'])
3 SaleQuantity=gb3.sort_values(by='销售数量',ascending=False)
4
5 TotalSales=SaleQuantity['销售数量'].sum()
6
7 # 计算百分比
8 SaleQuantity['Percentage']=SaleQuantity['销售数量']/TotalSales
9 SaleQuantity['AcmPerc'] = -1
10
11 # 计算累计百分比
12 for i in range(len(SaleQuantity.index)):
13     if i == 0:
14         SaleQuantity['AcmPerc'].iloc[0]=SaleQuantity['Percentage'].iloc[0]
15     else:
16         SaleQuantity['AcmPerc'].iloc[i]=SaleQuantity['Percentage'].iloc[i]+SaleQuantity['AcmPerc'].iloc[i-1]
17 Top_Sales=SaleQuantity[:10]
18
19 f,ax3=plt.subplots()
20 Top_Sales.plot(x=Top_Sales.index,y='销售数量',ax=ax3,kind='bar',title='销售前十',grid=True,rot=70)
21 ax4 = ax3.twinx()
22 Top_Sales.plot(x=Top_Sales.index,y='AcmPerc',kind='line',ax=ax4,secondary_y=True,marker='o',rot=70)
23 for a,b in zip(range(10),Top_Sales['AcmPerc']):
24     plt.text(a,b,'% .2f%%'%(b*100),ha='center',va='bottom')
25 ax3.legend(loc='upper center')

```



销售前十的药品如上图所示，且销售前十的药品销售数量占总销售数量的52%。

针对以上数据，还有很多可分析的内容，例如，是否有某药品集中在某一时期被购买的情况，哪些药品通常会被大规模购买，哪些药品在近一年内销售非常少可以适当减少采购量，需根据具体需求具体分析，在此就不再赘述。

我有个项目外包有人能做吗

小项目外包



想对作者说点什么



weixin_45254689： 作者您好，请能将这个练习的原始数据发给我吗？？ （1个月前 #4楼）



DylanGong： 求数据集，麻烦发到394863661@qq.com,谢谢！ （1个月前 #3楼）



m0_45173233： 😊😊，我也想要原始数据。 （1个月前 #2楼）

登录 查看 4 条热评

【python数据分析】——药品销售数据分析

阅读数 2884

一、前言这篇文章找来了有关药品销售数据进行案例分析练习，利用适当的统计方法对相关数据进行月...

博文 来自： 符智生的专栏

练习：药店销售数据分析

阅读数 2932

importpandaspdfilename='./朝阳医院2018年销售数据.xlsx'xls=pd.ExcelFile(filename,dtype='o...

博文 来自： 名为不二的兔...

第四关：pandas分析实际案例-药店销售数据分析

阅读数 584

importpandaspdfimportnumpyasnp#读取数据filename='E:\sale.xlsx'xls=pd.ExcelFile(filename,...

博文 来自： qq_42787271...

python数据分析案例——药店销售数据分析

阅读数 340

importpandaspdfilename='./朝阳医院2018年销售数据.xlsx'xls=pd.ExcelFile(fileNameStr,dty...

博文 来自： weixin_43139...

良心帖！看完这份路线，你的Python入门基础就差不多了！

Python学习路线免费领~

用Python实现数据化运营分析实例——销售预测

阅读数 2314

#案例场景：每个销售公司都有一定的促销费用，促销费用可以带来销量的显著提升；当给一定的促销...

博文 来自： Python_Big_lo...

数据分析项目实战：大型商场销售预测（带你挤进比赛前100名）

阅读数 8579

作者简介IntroductionJoffyZhong：咨询顾问|写作爱好者|数据分析|互联网创业者R语言中文社区专栏...

博文 来自： R语言中文社区



3



4

