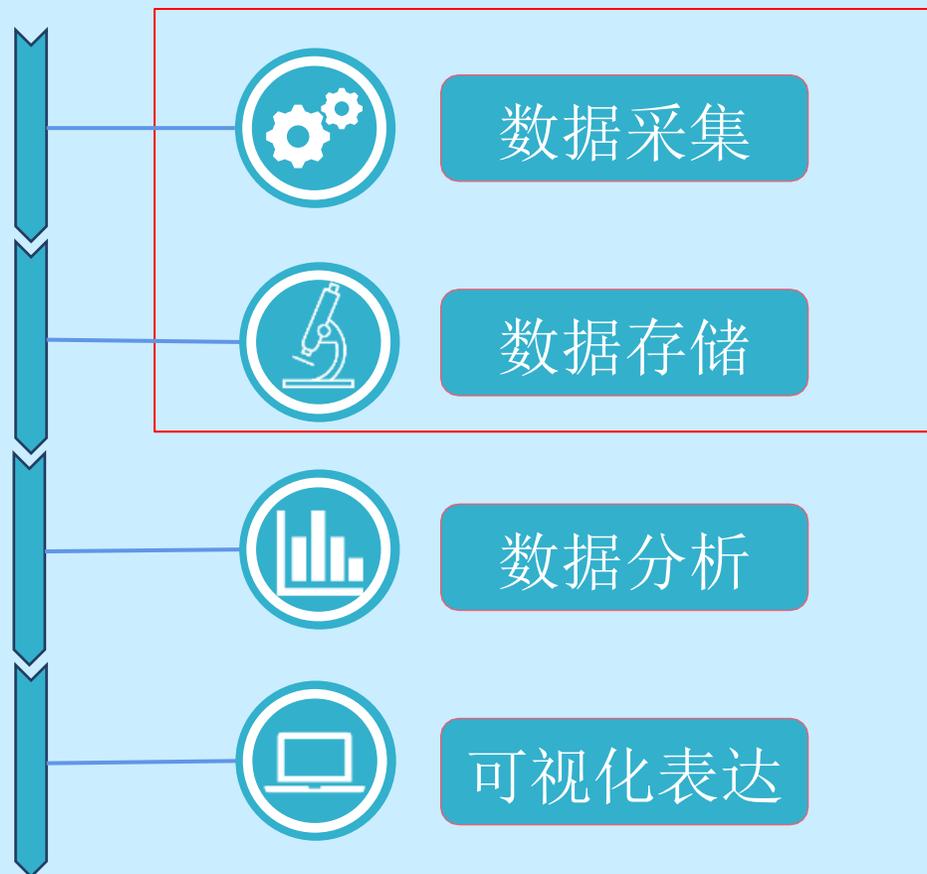




数据的分析（特征探索）

知识回顾



导入

某零食运营企业，需要调整运营策略。他们获取了近一段时间某零食购物网站的运营数据。

通过现有的数据，应该如何轻松并且准确得来探索他的特征，分析出对自己有价值的信息呢？



导入-数据展示

1	明治抹茶巧克力曲奇夹心饼干51%	喂你吃阿	49.80	74
2	手工糕点南瓜子饼营养食品瓜子饼干散装零食	返农归真	65.10	731
3	法国进口法芙娜黑巧克力62%萨蒂利亚黑巧	雅宏西餐	65.00	807
4	碧根果仁500g大山核桃仁长寿	贝乐滋旗舰店	23.50	248
5	马来进口饼干零食德散装	julies茱蒂丝旗舰店	24.90	54114
6	德国进口knoppers牛奶榛子巧克力2	天南帝北食品专营店	25.80	6237
7	粒鱼骨香酥骨头零食即食鱼深海	食味的初相旗舰店	25.90	167
8	蜂蜜黑芝麻香酥饼孕妇哺乳期月子零食薄脆饼	莹香轩旗舰店	24.90	793
9	意大利进口米兰大教堂珍藏版礼	zaini1913食品旗舰店	138.00	0
10	老板仔TAOKAENOI海苔6包29.9元	恒俞食品专营店	30.59	2541
11	Lotus和情缤咖时焦糖饼干125g4	和情旗舰店	24.90	903
12	澳门进口 乔凡娜 北海道巧克力奶油夹心爆	喂你吃阿	12.80	211
13	意大利进口黑金礼赞榛子	zaini1913食品旗舰店	118.00	22
14	闽南特产永春麻糍永盛珍猴麻糍散装	绘鑫特产小鑫	24.50	1434
15	歌斐颂85%黑巧克力纯可可脂纯黑苦巧克力	aficion歌斐颂旗舰店	79.00	3900
16	美荻斯进口坚果北美伯爵2800g干果礼盒	云集百汇	1699.00	0
17	歌斐颂58%纯黑巧克力纯可可脂散装批发网	aficion歌斐颂旗舰店	39.00	2938
18	澳大利亚进口TimTam雅乐思澳洲巧克力	研盛食品专营店	19.90	3636
19	Italyss坚果 天兴隆综合果仁每日坚	天兴隆旗舰店	24.90	1762
20	美国原装进口原味混合坚果1130g罐装	zjm张佳曼	115.00	7

数据分析的概念

数据分析是通过运用数字化工具和技术，对大批量杂乱无章的数据进行探索和分析，从中发现数据内在的结构和规律，并构建数学模型，进行可视化表达。



了解事物的现状



剖析事物的发展趋势



预测事物的未来走向

通过验证，将模型转化为知识，以便更好地了解事物的现状、剖析事物的发展趋势，并预测事物的未来走向。

探究新知

数据分析常用方法

1. 特征探索

2. 关联分析

3. 聚类分析

4. 建立模型

5. 模型评价



探究新知

数据特征探索

数据预处理

2	手工糕点南瓜子饼营养食品瓜子饼干散装零食	返农归真	65.10	731	
3	法国进口法芙娜黑巧克力62%萨蒂利亚黑巧	雅宏西餐	65.00	807	
4	贝乐滋碧根果仁500g大山核桃仁长寿	贝乐滋旗舰店	23.50	248	
5	马来进口饼干零食Julies/茱蒂丝雷蒙德散装	julies茱蒂丝旗舰店	24.90	54114	异常数据
6	德国进口knoppers牛奶榛子巧克力2	天南帝北食品专营店	25.80	6237	
7	粒鱼骨香酥骨头零食即食鱼深海	食味的初相旗舰店	25.90	167	
8	蜂蜜黑芝麻香酥饼孕妇哺乳期月子零食薄脆饼	莹香轩旗舰店		793	
9	Zaini 意大利进口米兰大教堂珍藏版礼	zaini1913食品旗舰店	138.00	0	
10	老板仔TACKAENCI海苔6包20.0元	恒合食品专营店	20.50	2541	

完成数据清洗

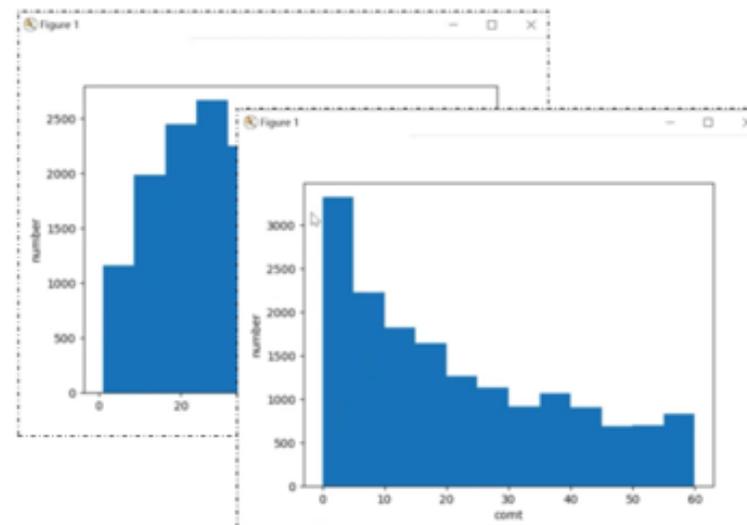
求最大值、最小值、极差绘制散点图，直方图

求最大值、最小值、极差、组距，绘制价格直方图和评论数直方图

```
#求最大值
pricemax=da2[2].max()
pricemin=da2[2].min()
commentmax=da2[3].max()
commentmin=da2[3].min()
#极差
pricerg=pricemax-pricemin
commentrg=commentmax-commentmin
#组距
pricedst=pricerg/13
commentdst=commentrg/13

#绘制价格直方图
#numpy.arange(最小,最大,组距)
pricesty=np.arange(pricemin,pricemax,pricedst)
plt.xlabel('price') #显示X坐标标签
plt.ylabel('number') #显示Y坐标标签
pyl.hist(da2[2],pricesty)
pyl.show()

#绘制评论数直方图
commentsty=np.arange(commentmin,commentmax,commentdst)
plt.xlabel('comt') #显示X坐标标签
plt.ylabel('number') #显示Y坐标标签
pyl.hist(da2[3],commentsty)
pyl.show()
```

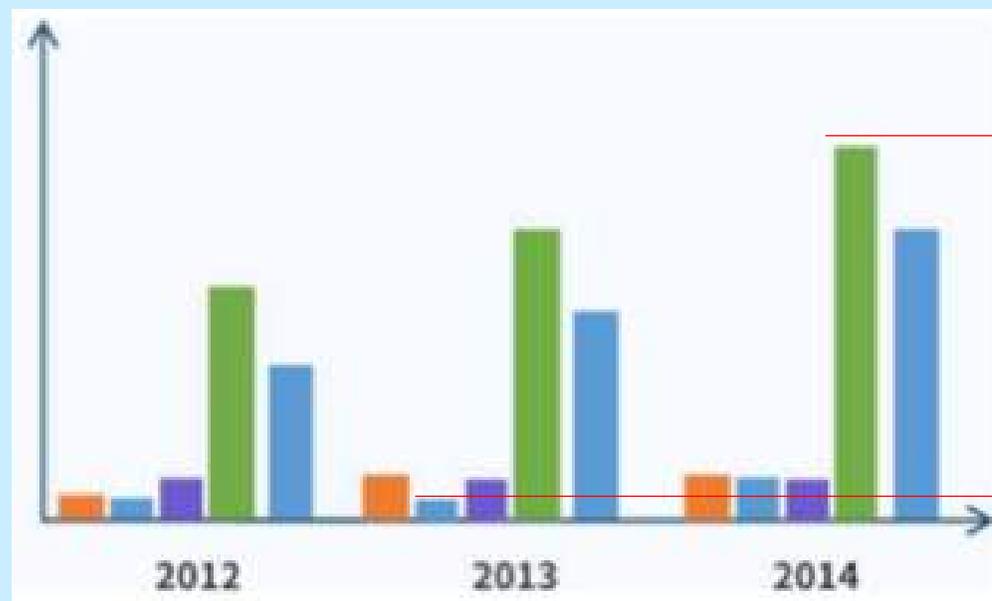


极差：

极差是最大值和最小值之间的差异或范围；

量化了数据的变异程度；

提供数据在整个取值范围内的大小差别。



①

绘制散点图，直方图



②

观察分析数据的分布特征

使用最大值、最小值、极差等描述性统计量。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/875040034244011314>