

Data Analysis and Working Experience II
数据分析在工作中的应用 2

Sen

提纲

- 提供准确，有效的数值 — 数据的收集和数据库维护
- 找出联系相关，因果相关 — 有效数据的收集和分析
- 给出预测，建议 — 数据的直观展示，计算和模拟

1.数据的收集 – 提供准确，有效的数值

1.1.现有数据

a. 现有数据库的日常维护

包括新增写入，分权限读取，数值更新，数据删除。

b. 建立维护的策略。

遇到空值，偏差，特殊值，例如重量的缺失。策略，1查厂家对照表，2近似型号的平均值，3实测存储的策略，根据数据大小，满足易操作 - Excel，速度优先的筛选 - Sql，易传播共享 - 服务器端安全和备份的策略，重要程度，Raid 1, Raid 10, 更多的分布式展示的策略，详尽的表，分布式的展示，客服部，采购部，物流部

c. 建立维护的日志。

本地备份，云端备份，版本的记录，精确到用户，部门的查看和修改记录

实际操作：

示例一，25000+ 商品数据的维护。包括名称，条码，货号，品牌，大小类别，属性，重量，单位，消费税，增值税，关税，跨境综合税，颜色，规格，等

*商品名称				*条码		*货号		*品牌名称		*大类名称		*子类名称	
Bellamy's Organic 贝拉米有机婴幼儿通心粉 8+200g				9332045000341		9332045000341		Bellamy 贝拉米		食品		食品	
属性名称	重量	单位	GST 消费税(%)	消费税率(%)	增值税(%)	关税(%)	跨境综合税(%)	*SKU 条形码	*SKU 编码	*颜色	*规格		
			0					9332045000341	9332045000341	默认	默认		

示例二，物流包裹数据的维护。包括寄、收件人姓名电话，省，城，区，邮编，身份证，商品，牌子，类别数量，规格，身份证，等

序号	交货时间(文)	运单号	寄件人	寄件人地址	寄件人电话	收件人	收件省	收件城市	收件区县	收件邮编	收件人地址	收件人电话		
207	2018-11-	207181129	Ga	澳洲	045170	霍					北京市顺	1355232		
207	2018-11-	207181129	Ga	澳洲	045170	王					内蒙古呼	1584760		
207	2018-11-	207181129	小	澳洲	838572	王					北京顺义	1381012		
207	2018-11-	207181129	Ter	澳洲	047852	刘					山西省太	1319108		
计费重量(磅)	实际重量(磅)	数量(个)	运费(USD)	保价金额(USD)	保额(USD)	品名	品牌	类别	规格	单价(RMB)	物品数量	总申报价值(USD)	渠道	身份证号码
3						DK	37DK	鞋子			1			
6						DK	36DK	鞋子			1			
0						DK	38DK	鞋子			1			
9						DK	39DK	鞋子			1			

1.2.网上截取

页面数据的爬虫收集

- 数据源于平台，对自己商品的定价，名称的优化，SEO
- 数据源于潜在竞争对手网站
- 其他媒体的播放量，及社会热度的关注

- 数据源于平台，对自己商品的定价，名称的优化，SEO

5月27日 京东 搜 iphone 前 213 条数据，关键词1. 苹果 2.手机型号的列举 3. 认证

标题	标题链接	价格	promo-words	p-commit	销量
绿联 MFi认证 苹果数据线Xs Max/XR/X/8	https://item.j	39.90	【热卖】合金机身，硅胶材质，防	二手有售76万+条评价	76万+
爱心东东无线蓝牙耳机iPhone/XS/Max/XR	https://item.j	238.00	玩游戏不延迟不卡顿，HIFI音效	6500+条评价	6500+
正件 苹果车载音频线aux插口链接线 iphon	https://item.j	49.00	【京东自营】镀金接头即插即用	200+条评价	200+
网易严选 网易智造 MFi认证 苹果数据线 Xs	https://item.j	39.90	苹果官方MFi认证，品质保证！闪	二手有售47万+条评价	47万+

b. 数据源于潜在竞争对手网站，竞品类别，定价

三年前 今日悉尼 会员商城 489 条数据，1 其他公司经营的类别，成人婴幼儿保健品奶粉， 2 和品牌的合作，

SKU	品牌	品类	中文产品名	英文产品名
9347832000008	Wyeth / 惠氏	婴幼儿奶粉	惠氏金装S26婴幼儿奶粉1段	S26 NEWBORN STAGE 1
9322635005436	Wealthy Health / 富康	普通药物	富康有机角鲨烯 365粒	Wealthy Health Organic Squa
9300652010794	WEET-BIX	成人健康	WEET-BIX全麦高纤有机麦片	SANITARIUM WEET-BIX
5000174471748	Vicks Vaporub	婴幼儿健康	Vicks Vaporub 婴儿通鼻薄荷软膏 50g	Vicks Vaporub Baby Balsam
9350634002859	Unichi	成人健康	Unichi玫瑰果精华胶囊 美白 修复 抗黑色素	Unichi Rosehip Extract Comp
9350634002842	Unichi	成人健康	unichi 葡萄籽精华 美白抗氧化素抗衰老 60粒	Unichi Tannat Grape Seed™

c. 其他媒体的播放量，及社会关注的关注度

5月27日 SBS Dateline 的YouTube 频道 最近112 个视频

同样一个月

香港的贫富悬殊 193 万播放量

How Hong Kong is home to the crazy rich and the mega poor

毛里塔尼亚 的女孩子，被强制吃多点，看起来成熟有钱，

3000播放量，Young girls in Mauritania are force fed to look older and wealthier for marriage Doc

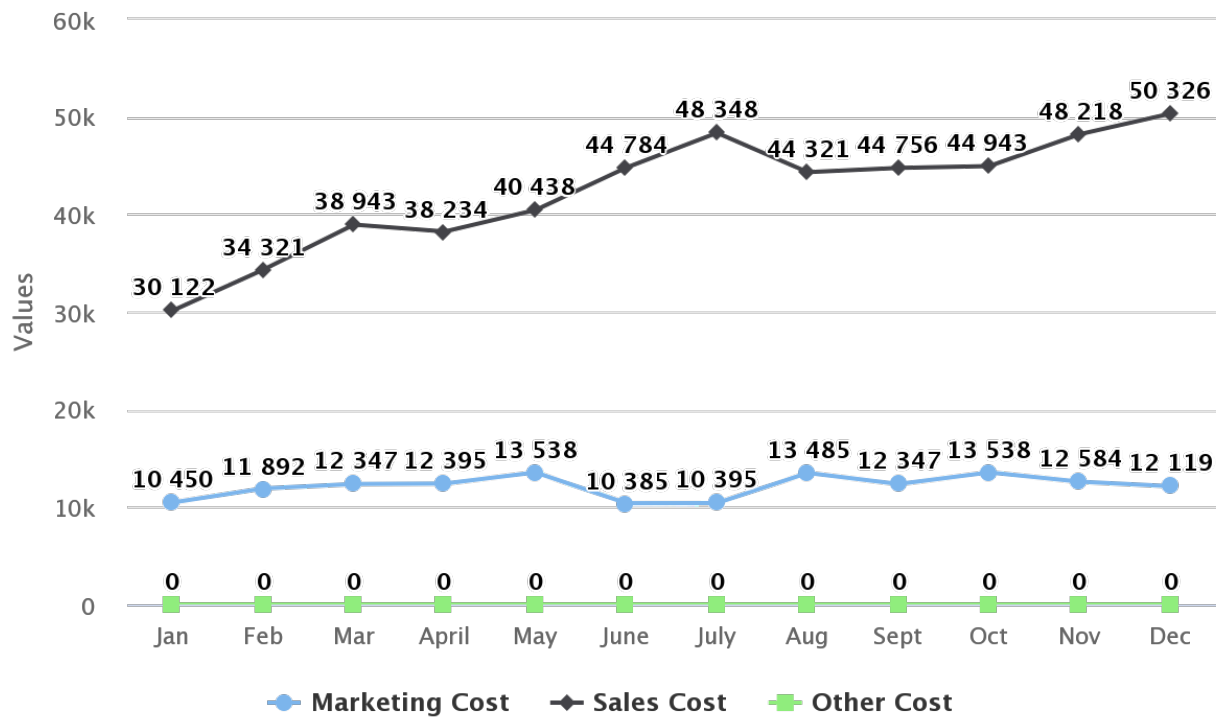
标题	标题链接	style-scope1	style-scope	缩略图
The Children of Agent Orange	https://www.you	2530000	1年前	https://i.yt
How Hong Kong is home to the crazy rich and the mega poor	https://www.you	1930000	1 个月前	https://i.yt
Why are Filipino communities living in cemeteries and caring for its dead?	https://www.you	1270000	1年前	
Why are Indigenous women missing in Canada?	https://www.you	980000	1年前	
Why are young New Zealand Māori barely holding on?	https://www.you	960000	8 个月前	https://i.yt
Inside the Brazilian prison run by the inmates	https://www.you	910000	7 个月前	https://i.yt
What happens to the children left behind by China's industrial boom?	https://www.you	870000	1年前	https://i.yt
The man building tiny homes for the homeless in Los Angeles	https://www.you	820000	1年前	

Google Analytic 对网站的监控，手机端，桌面端

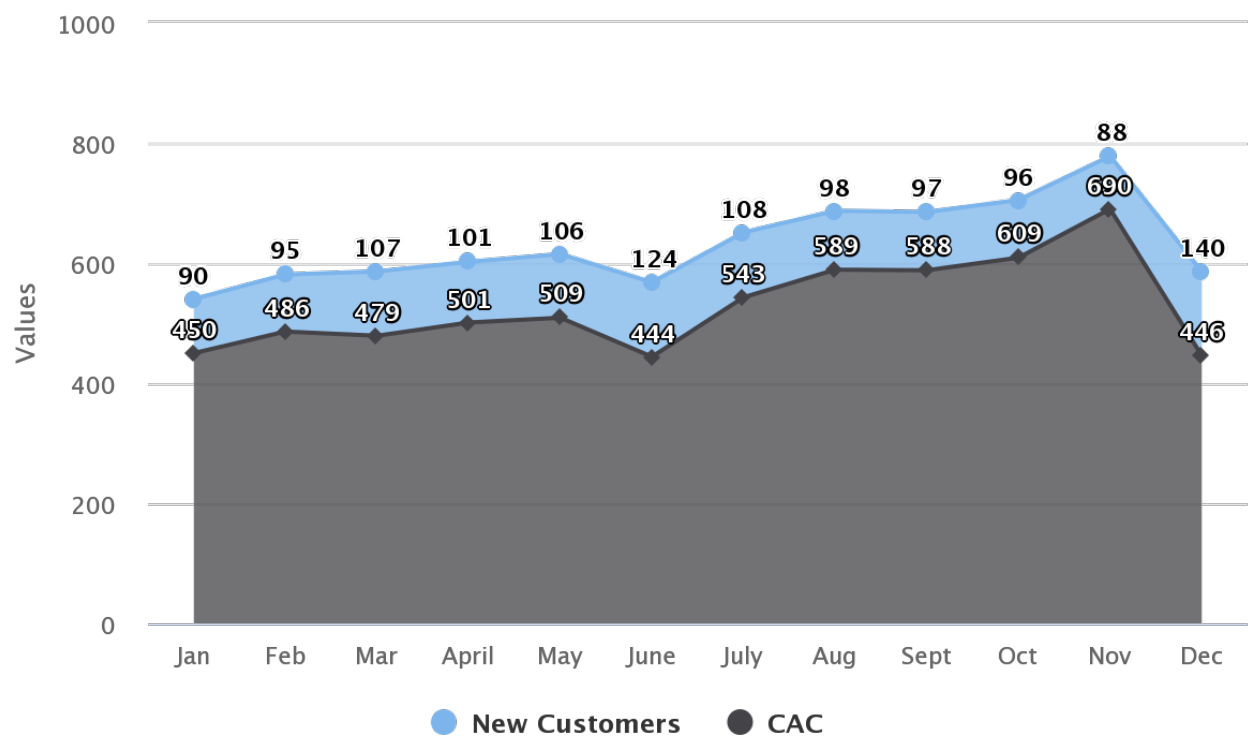


2.数据的可视化

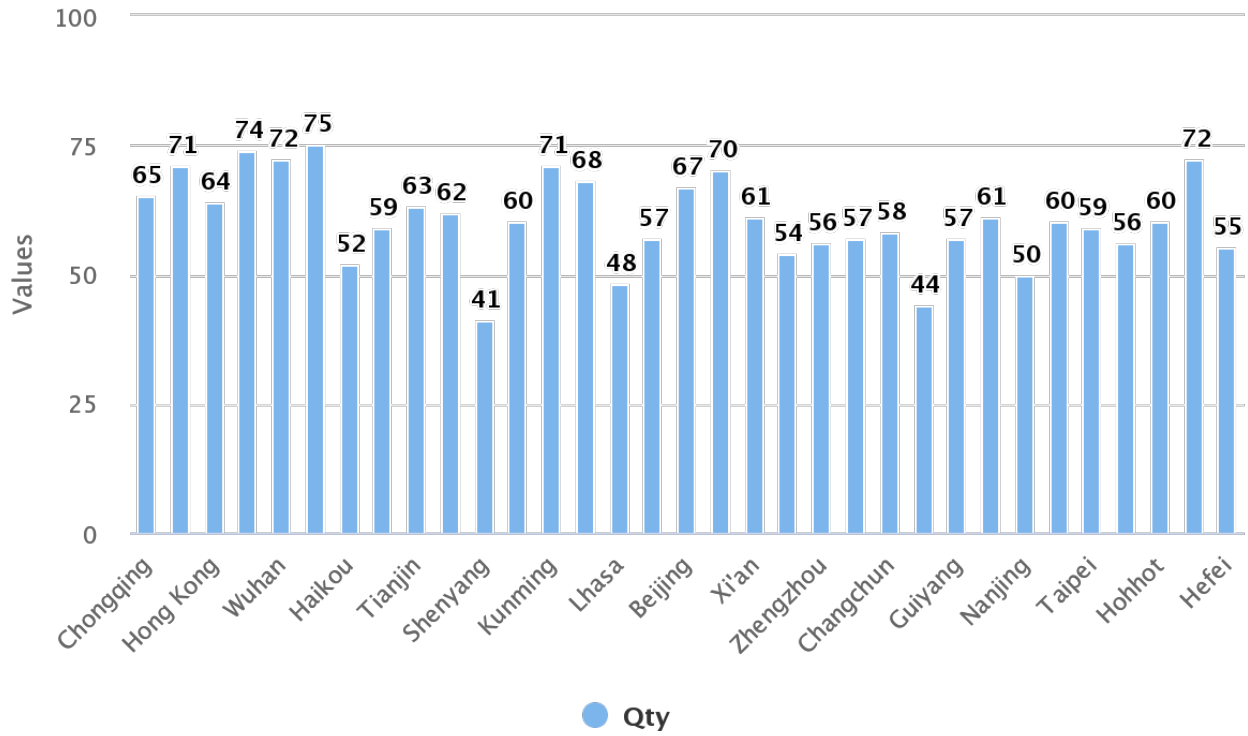
Cost Comparison



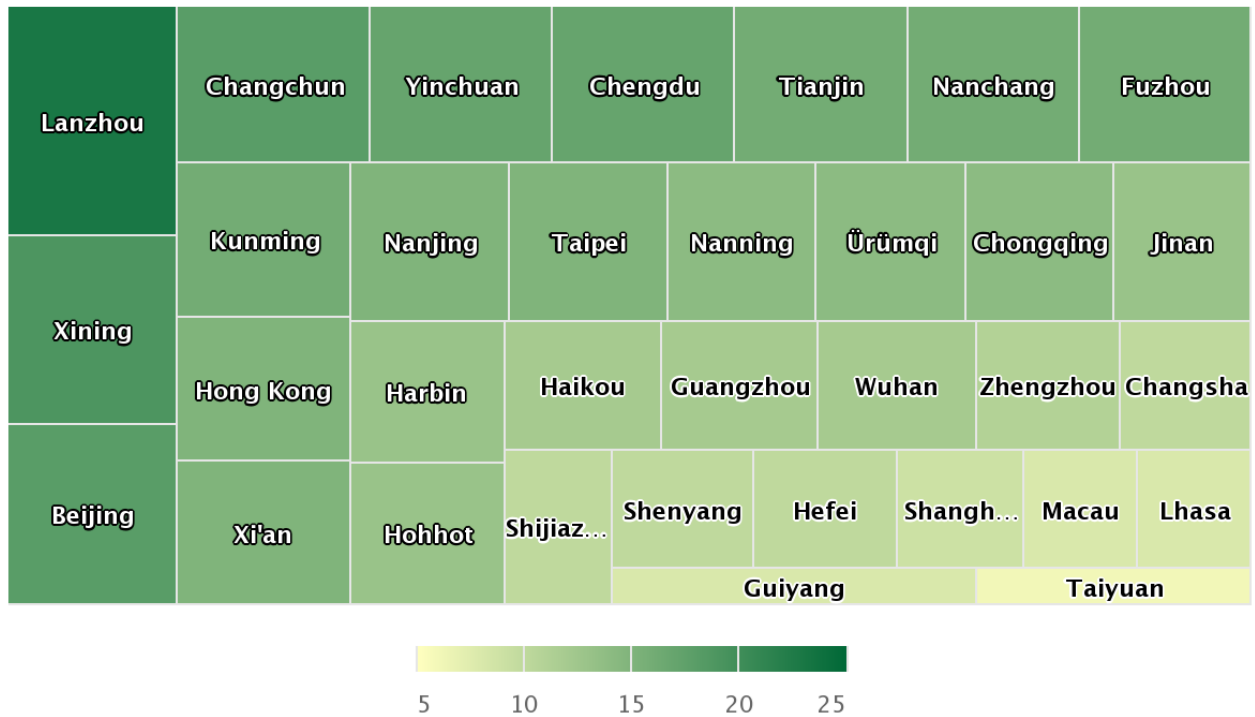
Customer Growth



Total Sales Volume of each City



Ozlana Sales in different Cities



Quick View — 物联网 — Capstone Project

	传感器 ->	单片机 ->	云服务器 ->	手机 App ->
数据流	采集	监控	存储 + 展示	展示 + 远程控制
工具	Python + Linux Shell		API 接口	iOS, Swift
示例	温度传感器检测 温度大于 45 度	单片机自动控制 打开风扇	记录每 20 秒的数据点 折线图展示一段时间内所有数值	显示当前温度 45 远程控制所有开关

