



CDA 数据分析师
CERTIFIED DATA ANALYST

CDA 企业内训通信行业热门课纲

一、业务分析师（一级）内训课纲（9 天）	2
二、Python 数据挖掘算法课纲（二级建模）（6 天）	7
三、机器学习课纲（18 天）	9
四、大数据分析应用课纲（5 天）	11
五、Power BI 培训（2 天）	13
六、Python 基础及可视化（2 天）	14
七、大数据认知与应用（4 天）	20
八、大数据治理实战（4.5 天）	25
九、如何撰写有价值的数据分析报告（1-2 天）	28



一、业务分析师（一级）内训课纲（9天）

课程模块	内容设置	课时
概率论与数理统计基础	<p>part1: 数据分析方法概述</p> <ol style="list-style-type: none"> 1、数据分析过程 2、数据分析的商业驱动 <p>part2: 概率论基础</p> <ol style="list-style-type: none"> 1、随机事件及其概率 2、概率的性质与运算法则 3、离散型随机变量及其分布 4、连续型随机变量的概率分布 <p>part3: 描述性统计</p> <ol style="list-style-type: none"> 1、集中趋势的度量 2、离散程度的度量 3、偏态与峰态的度量 <p>part4: 常见分布族</p> <ol style="list-style-type: none"> 1、分布概念 2、正态分布及其衍生重要分布 3、中心极限定理 4、样本比例抽样分布 <p>part5: 多维随机变量</p> <ol style="list-style-type: none"> 1、联合分布 2、协方差 3、相关系数 <p>part6: 数据简化原理</p> <ol style="list-style-type: none"> 1、似然函数 2、辅助函数 <p>part7: 参数估计</p> <ol style="list-style-type: none"> 1、点估计 2、区间估计 	2天

	<p>3、样本量的确定</p> <p>part8: 假设检验</p> <p>1、单侧检验</p> <p>2、总体参数检验</p> <p>3、P 值</p> <p>part9: 方差分析</p> <p>1、单因素方差分析</p> <p>2、双因素方差分析</p> <p>part10: 一元线性回归</p> <p>1、变量间关系的度量</p> <p>2、一元线性回归方程</p> <p>2、线性相关检验</p> <p>3、残差分析</p>	
<p>Mysql 数据库管理</p>	<p>part0: MySQL 的安装与卸载</p> <p>1、MySQL 安装</p> <p>2、MySQL 的报错解决</p> <p>3、MySQL 环境变量配置</p> <p>part1: 数据库基本操作</p> <p>1、MySQL 数据库的创建：创建、查看、使用、删除数据库</p> <p>part2: 数据表操作</p> <p>1、数据表与数据库之间的关系</p> <p>2、数据表的创建、查看、删除语句</p> <p>part3: 数据库、数据表与字段操作</p> <p>1、数据类型：整数型、小数型、日期和时间、字符串</p> <p>2、约束条件：主键约束、非空约束、唯一性约束、默认约束、自增字段</p> <p>part4: SQL 查询与函数</p> <p>1、查询语句：单表查询、多表查询、内连接查询、左连接、右连接、聚合查询等</p>	<p>2 天</p>

	<p>2、聚合函数：AVG、SUM、MAX、MIN、COUNT</p> <p>3、查询操作符与子查询：and、or、in、like、is null、distinct</p> <p>4、操作符与子查询的组合应用</p> <p>5、as 重命名与 limit 限制查询结果行数</p> <p>6、函数：常用数学函数、字符串函数、日期及时间函数、其他函数</p> <p>7、为字段赋值</p> <p>8、删除记录</p> <p>part5: SQL 查询应用</p> <p>1、电商数据表 ER 图</p> <p>2、电商数据多表查询练习</p>	
<p>SPSS 数据分析实战 (SPSS 为实现工具， 主要讲解案例)</p>	<p>part0: SPSS 软件介绍</p> <p>1、数据分析流程介绍：以员工短期绩效统计为例</p> <p>2、SPSS 特点</p> <p>3、SPSS 界面及操作介绍</p> <p>part1: 数据的输入与保存</p> <p>1、数据的获取与数据格式</p> <p>2、数据变量详解</p> <p>3、SPSS 访问外部数据</p> <p>part2: 数据的预处理</p> <p>1、数据清洗</p> <p>2、变量衍生</p> <p>3、描述性统计分析功能</p> <p>4、使用 SPSS 绘制常用统计图形</p> <p>part3: 一元线性回归</p> <p>1、最小二乘估计</p> <p>2、线性回归与相关</p> <p>3、线性回归与方差</p> <p>4、数据分析流程</p> <p>part4: 多元线性回归</p>	<p>4 天</p>

<p>1、正态多元线性回归的假设</p> <p>2、正态分布的问题</p> <p>3、异方差问题与处理</p> <p>4、异常值问题与处理</p> <p>5、共线性问题与处理</p> <p>part5: Logistic 回归</p> <p>1、logistic 回归与卡方：卡方分析原理、相关系数指标（kappa、风险、OR 等）</p> <p>2、最大似估计</p> <p>3、logistic 回归解析：模型与线性回归对比、两种 R2 对比、回归系数与 OR 值、RR 值、OR 值域 gamma 值</p> <p>4、评分与预测：老样本预测与市场细分、数据的分箱化、数据离散化、评分卡模型中观点的意义</p> <p>part6: 特征筛选、主成分分析</p> <p>1、特征筛选：特征筛选流程、DB 特征筛选合理性、DB 特征筛选方法</p> <p>2、主成分回归：主成分原理、主成分判断标准、主成分的应用场景、主成分的因子、主成分回归</p> <p>part7: 市场细分-客户画像</p> <p>1、用户画像应用：商业场景、模型算法应用</p> <p>2、模型变量的确定：标签数、标签筛选</p> <p>3、聚类算法：K-Means 聚类、两步聚类、Kohonen 聚类</p> <p>4、市场细分：分组</p> <p>5、标签制作原则</p> <p>part8: 市场细分-客户价值模型</p> <p>1、FRM 模型：商业应用、模型算法应用</p> <p>2、因变量的确定：理论性及专家意见、潜变量及显变量评价</p> <p>3、价值细分：分箱化、秩变换</p> <p>4、价值细分应用：价值细分、价值评分</p> <p>part9: 时间序列</p>	
---	--

	<p>1、时间序列原理介绍：AR 模型、MA 模型、ARIMA 模型</p> <p>2、时间序列数据预处理：定义时间变量、时序数据的处理、序列图形</p> <p>3、使劲序列的建模与预测：确定性分析、随机性分析</p> <p>part10：不同市场订户信息的序列预测</p>	
考点讲解及测试	<p>Level 1 考点讲解</p> <p>模拟测试</p>	1 天

二、Python 数据挖掘算法课纲（二级建模）（6 天）

01 章 Python 数据挖掘基础理论与数据预处理技术

01-01 数据挖掘概要

01-02 数据挖掘的方法和原理

01-03 数据挖掘基础和进阶技术概述

01-04 数据预处理技术：字段选择-数据清洗-字段扩充-数据编码

01-05 人工特征工程：特征构造 - 特征抽取 - 特征选择

02 章 Python 数据挖掘之决策树算法

02-01 决策树建模思路

02-02 Quinlan 系列决策树 (ID3、C4.5、C8.0) 建模原理

02-03 CART 建模原理

02-04 决策树模型修剪

02-05 决策树模型效果评估

02-06 案例：使用决策树进行初始信用评级

03 章 Python 数据挖掘之神经网络和贝叶斯网络

03-01 了解神经网络的基本概念

03-02 明确人工神经网络结构

03-03 神经元模型

03-04 掌握 BP 神经网络学习算法

03-05 多层感知器的 scikit-learn 代码实现

03-06 贝叶斯公式与分类原理

03-07 朴素贝叶斯的参数估计

03-08 在 Python 中实现朴素贝叶斯

04 章 Python 数据挖掘之分类器与集成学习方法

04-09 KNN 算法原理与 Python 实现

04-01 线性可分与线性不可分

04-02 线性可分的支持向量机

04-03 线性支持向量机与软间隔最大化

04-04 非线性支持向量机与核函数

04-05 集成学习方法: Bagging、Boosting、随机森林

05 章 Python 数据挖掘之聚类分析

05-01 聚类算法的概述

05-02 聚类算法基本概念

05-03 聚类模型的评估

05-04 层次聚类原理与 Python 实现

05-05 基于划分的聚类 K-means 的原理及应用

05-06 详谈基于密度的聚类方法与 Python 实现

05-07 案例: 通信客户业务使用偏好聚类

06 章 Python 数据挖掘之关联规则与序列模式

06-01 关联规则的一些基本概念

06-02 关联规则 Apriori 算法的原理与 Python 实现

06-03 关联规则 FP-growth 算法

06-04 序列模式的简介与概念

06-05 序列模式 AprioriAll 算法与 Python 实现

06-06 基于用户和商品的协同过滤算法仅供参考



三、机器学习课纲（18 天）

大类	小类	培训项目	说明	时间 (天)
建模基础	数学基础	高等数学	高等数据基础与微积分	2
		概率论	概率基础、概率分布、最大似然估计等	
		统计学	描述统计、推断统计	
		线性代数	向量、行列式、矩阵、特征值与特征变量、奇异值分解等	
		最优化	梯度下降法、牛顿法等	
	PYTHON基础	基础知识	安装配置、基础语法等	4
		常用库	Numpy、Pandas、Matplotlib、Ipython、	
		机器学习库	Scipy、Scikit-learn、TensorFlow	
		数据处理	数据清洗、数据标准化方法、缺失值处理等	
		模型评价	模型评价的方法，如准确率、召回率、KS、PSI 等	
机器学习算法	分类算法	逻辑回归算法	算法原理、Python 实现、示例讲解	6
		决策树算法	算法原理、Python 实现、示例讲解	
		K 近邻算法	算法原理、Python 实现、示例讲解	
		朴素贝叶斯算法	算法原理、Python 实现、示例讲解	
	回归算法	线性回归算法	算法原理、Python 实现、示例讲解	
		岭回归算法	算法原理、Python 实现、示例讲解	
		Losso 回归算法	算法原理、Python 实现、示例讲解	
	聚类算法	K-Means 算法（基于划分）	算法原理、Python 实现、示例讲解	
		DBSCAN 算法（基于密度）	算法原理、Python 实现、示例讲解	
		层次聚类法	算法原理、Python 实现、示例讲解	
	降维算法	主成分分析算法（PCA）	算法原理、Python 实现、示例讲解	
		线性判别分析算法（LDA）	算法原理、Python 实现、示例讲解	
		LLE 局部线性嵌入算法（非线性降维）	算法原理、Python 实现、示例讲解	

	集成学习算法	Bagging/Boosting	算法原理、Python 实现、示例讲解	2
		GBDT 算法	算法原理、Python 实现、示例讲解	
		随机森林算法	算法原理、Python 实现、示例讲解	
		Adaboost 算法	算法原理、Python 实现、示例讲解	
		Xgboost 算法	算法原理、Python 实现、示例讲解	
	深度学习算法	文本挖掘算法	算法原理、Python 实现、示例讲解	2
		关系图谱算法	算法原理、Python 实现、示例讲解	
	其他算法	神经网络算法	算法原理、Python 实现、示例讲解	
		关联规则算法	算法原理、Python 实现、示例讲解	
		SVM 支持向量机算法	算法原理、Python 实现、示例讲解	
实践项目	案例	客户生命周期案例	结合客户生命周期案例，讲解精准营销、银行反欺诈、信用评分等内容。	2

四、大数据分析应用课纲（5天）

时间	培训主题	培训内容
第一天上午	大数据前沿知识 及 Hadoop	<ol style="list-style-type: none"> 1. 大数据前沿知识介绍 2. 课程介绍 3. Linux 及 Ubuntu 安装和使用 4. Linux/Ubuntu 文件系统操作 5. Hadoop 的单机、伪分布 <p>【操作】 完成 Hadoop 伪分布搭建</p>
第一天下午	Hadoop 集群搭建 及 HDFS 操作	<ol style="list-style-type: none"> 1. Hadoop 完全分布式模式的安装配置 2. Hadoop2.x 体系结构 3. HDFS 分布式文件系统，HDFS Shell 操作 4. YARN 的基本构成和工作原理 <p>【操作】 完成 Hadoop 集群搭建</p>
第二天上午	Mapreduce 思想及案例	<ol style="list-style-type: none"> 1. MapReduce 并行计算框架 2. 使用 MapReduce 编程框架实现数据去重 3. 使用 MapReduce 编程框架实现数据排序 3. Hadoop 集群上部署和执行 MR Job <p>【案例】: 论坛日志分析</p>
第二天下午	大数据导入与存储 Tableau 大数据分析	<ol style="list-style-type: none"> 1. mysql 数据库基础知识 2. hive 的基本语法 3. hive 安装部署 4. Tableau 的维度与度量 5. 使用 Tableau 快速创建交互式绘图 6. 利用 Tableau 的 GUI 创建交互式仪表盘 7. 使用 Tableau 连接 mysql 数据源 <p>【操作】: Tableau 实践</p>
第三天上午	Hbase 理论及实战	<ol style="list-style-type: none"> 1. Hbase 简介、安装及配置 2. Hbase 的数据存储

		<p>3. Hbase Shell Hbase API</p> <p>4. Hbase 数据备份与恢复方法</p> <p>【案例】: 针对 XX 高校学生管理信息系统中的学生选课</p>
第三天下午	Spark 配置及使用场景	<p>1. Scala 基本语法</p> <p>2. Spark 介绍及发展历史</p> <p>3. Spark standalone 模式部署</p> <p>4. Spark RDD 详解</p> <p>5. Spark 案例分析</p> <p>【操作】: Spark1.4 Standalone 分布式集群</p>
第四天全天	Spark 大数据分析原理	<p>1. Spark 作业调度流程与策略</p> <p>2. 集群模式运行、监控 Spark 应用程序的步骤与方法</p> <p>3. Hive、Shark 简介与对比分析</p> <p>4. Spark MLlib 体系结构简介</p> <p>5. 机器学习简介及经典算法案例详解</p> <p>6. Spark SQL 体系简介及语句流程</p> <p>7. DataFrame 模型的定义、功能</p> <p>【案例】: Sogou 搜索日志数据之用户行为进行分析</p>
第五天上午	Hadoop+Spark 智慧高速大数据分析	<p>1. 复习 Hive、Sqoop、Spark、Mysql 的配置</p> <p>2. Sqoop 和 Mysql 数据库进行交互</p> <p>3. Spark SQL 语句操作大数据平台上的数据</p> <p>4. Tableau 进行高速大数据的漏斗分析</p>
第五天下午	Hadoop+Spark 互联网金融监管大数据分析	<p>1. 使用爬虫进行文本数据采集</p> <p>2. 使用 spark 实现文本分词</p> <p>3. 互联网金融公司提取</p> <p>4. 公司名词判别</p> <p>5. 使用 SPARK 程序实现情感语义分析</p> <p>6. 结果展示</p>

五、Power BI 培训（2 天）

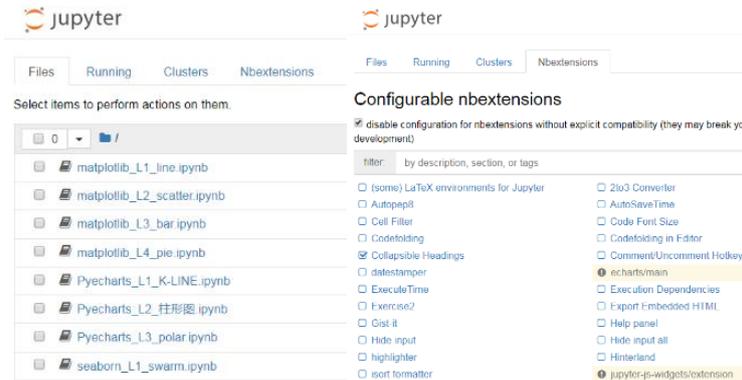
模块	内容	时长
Excel 进阶 Power BI 商业智能	<ol style="list-style-type: none"> 1. Power BI 简介与数据可视化 2. Power Query 基本功能：数据导入、数据转换、数据组合、加载数据模式、合并查询等 3. Power Query 常用移除文本操作 4. Power Pivot 数据导入功能 5. Power Pivot 搭建多维数据集与多维数据透视 6. Power Pivot 基本功能：数据加载、计算字段、KPI、层次结构等 Power Pivot 综合应用案例：商机相关企业信息	6h
客户数据分析案例	<ol style="list-style-type: none"> 1. 经典客户分析方法介绍 2. RFM 分析：“最近 12 个月”动态客户 RFM 分析 3. 帕累托分析：多种方法构造的交互式帕累托分析 4. 客户价值分层分析 5. 客户 ABC 分析方法 	3h
销售管理数据分析 案例	销售部业务分析仪表盘报告 <ol style="list-style-type: none"> 1. 销售管理分析-销售漏斗分析方法 2. 销售运营业务图 3. 销售管理分析 - 搭建多维数据模型：客户维度、产品维度、商机维度、地域维度、渠道维度、销售维度 4. 销售管理分析-洞察风险 5. 销售管理分析设计思路 6. Power Query 进行数据加工 7. Power Pivot 进行数据加工 8. 仪表盘展现总体趋势问题 9. 瀑布图展现各阶段占比情况 10. 了解各个阶段商机构成 11. 销售管理分析仪呈现与细节问题 	3h

六、Python 基础及可视化（2 天）

第一天

(1) 厉兵秣马——Python 简介与 Jupyter 配置

样例：



内容简介：学习基本的环境设置，为后期的代码编辑做准备。

(2) 雾里看花——可视化基本逻辑与商业应用

内容简介：

- 数据可视化是什么？
- 数据可视化如何在不同的商业场景中进行应用？
- 当下热门的“一图看天下”是什么？
- 数据可视化的误区有哪些？

(3) 无中生有——变量的理解与随机数据生成

样例：

构造样本数据1

```
In [2]: num=100
np.random.seed(1)
capital=np.random.randn(num)*10+100+np.random.rand(num)*20
np.random.seed(2)
sale=1.2*capital+20+np.random.randn(num)*20
np.random.seed(3)
cost=-0.8*capital+200+np.random.randn(num)*5

In [3]: data=pd.DataFrame([capital, sale, cost]).T
data.columns=['capital', 'sale', 'cost']
print(data.head())
```

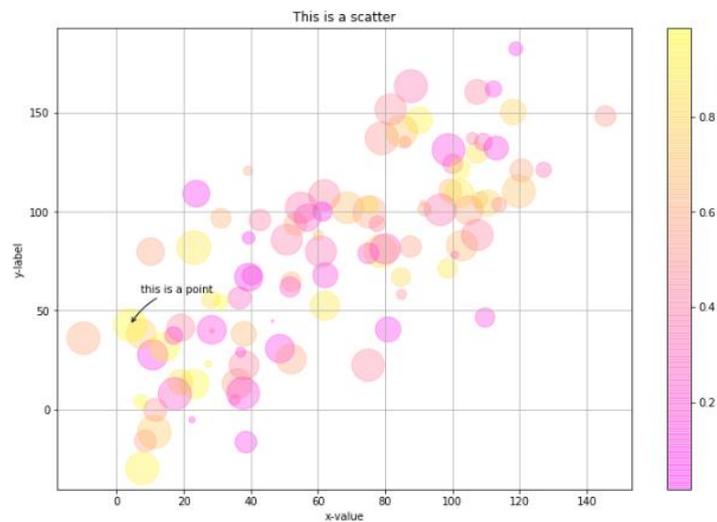
	capital	sale	cost
0	132.385558	170.527512	103.034696
1	101.639649	140.842242	120.870830
2	111.989120	111.663022	110.891192
3	104.212747	177.860712	107.312339
4	119.778881	127.865945	102.789954

内容简介:

- 产生模拟的数据
- 对数据进行拼接
- 对数据进行分割
- 数据的筛选与补充

(4) 循序渐进——散点图与条形图的进化过程

样例:

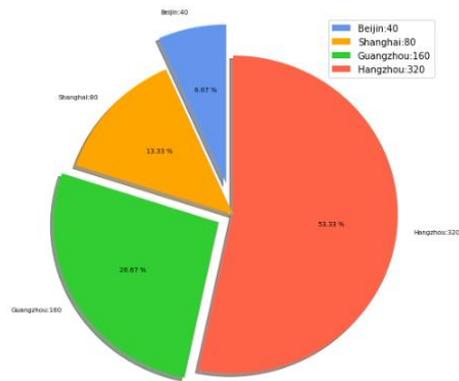


内容简介:

- 散点图和条形图的适用数据类型
- 散点图的主要参数
- 条形图的主要参数
- 散点图与条形图的实现
- 可视化过程中的数据降维问题

(5) 渐入佳境——用饼图与环形图呈现数据比例

样例:

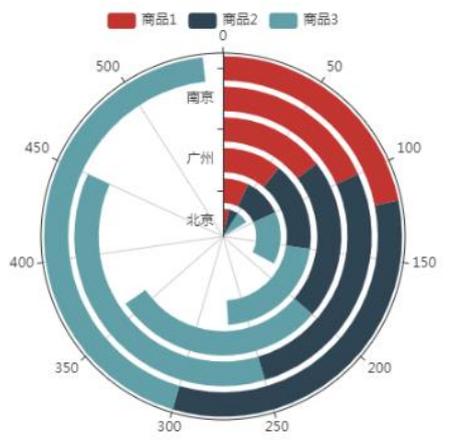


内容简介:

- 饼图与环形图的适用场景
- 饼图与环形图的关联与区别
- 饼图与环形图的重要参数
- 饼图与环形图的代码实现

(6) 秘境寻踪——南丁格尔的往事与玫瑰花图

样例:



内容简介:

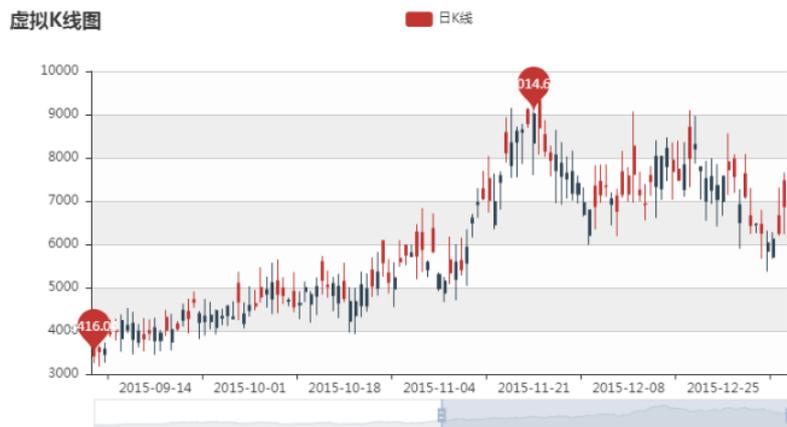
- 南丁格尔与玫瑰花图的起源
- 玫瑰花图的基本原理
- 玫瑰花图的衍生
- 相关图像的代码实现

第一天学员收获: 学员掌握 Python 基本的数据赋值, 数据分解与合并, 数据构造, 变量转换, 条件语句, 循环语句等。同时能够运用 Python 进行可视化编程。

第二天

(1) 有条不紊——折线图的设计与多线条呈现

样例：

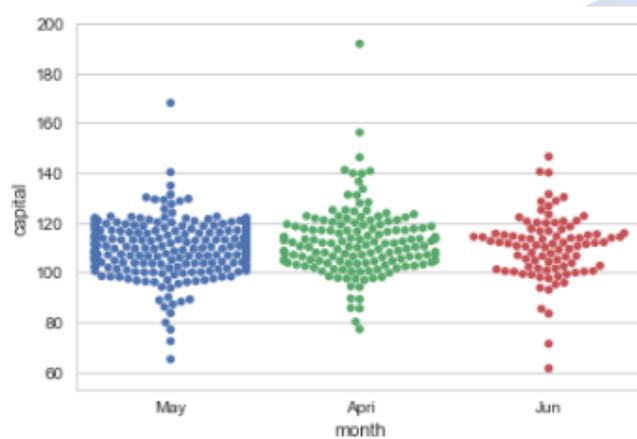


内容简介：

- 折线图的应用范围
- 折线图的主要参数
- 折线图的代码实现
- 多折线的呈现技巧

(2) 火眼金睛——分布与箱线图的异常值发现

样例：

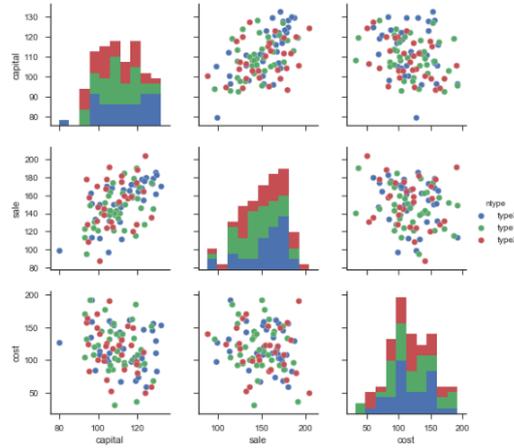


内容简介：

- 什么是数据分布
- 数据分布与异常值
- 分布与箱线图的主要参数
- 分布与箱线图的代码实现

(3) 抽丝剥茧——关联与配对图的相关性识别

样例：

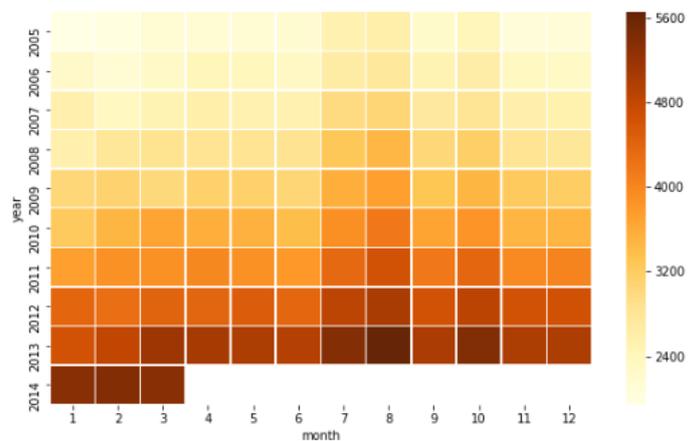


内容简介：

- 关联图与配对图的关系与区别
- 什么是相关性
- 关联图与配对图的主要参数
- 关联图与配对图的代码实现
- 图像内涵的信息解释

(4) 冰火相融——日历热力图中的周期性探索

样例：



内容简介：

- 热图的使用场景
- 热图的数据要求

- 热图的核心参数
- 热图的代码实现
- 如何解读热图

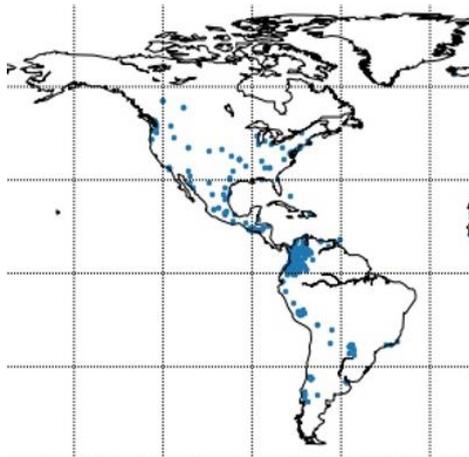
(5) 妙笔生花——可视化数据的动画设计基础

内容简介:

- 动画设计的基本原理
- 动画设计的误区
- 动画设计如何结合具体问题
- 动画的代码实现

(6) 图看天下——基于地理信息的数据可视化

样例:



内容简介:

- 什么是地理信息可视化
- 地理信息可视化的注意要点
- 地理信息可视化的数据要求
- 地理信息可视化的参数
- 地理信息可视化的代码实现与解读

第二天学员收获: 学员较为熟练运用 Python 中的 Numpy, Pandas 等常用库。能够在不同的具体场景选择不同的可视化方案并实现。

七、大数据认知与应用（4天）

Part 1 数据化思维导论

1 走近数据化管理

1.1 什么是数据化管理（30 min）

1.1.1 数据化管理的四个层次

1.1.2 驱动业务的三驾马车

1.1.3 经验与创新

1.2 数据思维（30min）

1.2.1 立体化：点、线、面

1.2.2 元素化：时间、对象、事件

1.2.3 多角度：广度、宽度、深度

1.2.4 常见误区解析

1.3 增长的秘密（60 min）

1.3.1 从0到1：新业务线的爆发式推进

1.3.2 从1到N：消费升级与品牌价值

1.3.3 新零售：融合流量与场景的生态圈

1.4 数据驱动（60 min）

（企业维度）

1.4.1 品牌价值评估

1.4.2 业务线分析

1.4.3 供应链管理

（客户维度）

1.4.4 客户生命周期管理

1.4.5 大数据营销

1.4.6 客户行为分析

1.5 常见问题与困难（60min）

1.5.1 如何将问题抽象成数据的语言：概率与分布

1.5.2 如何选择恰当的维度：方差分析与群体切分

1.5.3 如何设计实验方案：对照组与实验组的选择

- 1.5.4 如何评估最终效果：常用指标设定与调整
- 1.5.5 缺失数据怎么处理：常用填补方法
- 1.5.6 结果不符合预期怎么处理：寻因法与纠错法
- 1.5.7 陌生问题没有思路怎么处理：矩阵式分析与图形化思考
- 1.5.8 考虑的方面过多怎么处理：降维与独立变量法

2 走近客户：想说懂你不容易

2.1 了解客户：特征画像（60min）

- 2.1.1 为什么我们需要做客户画像
- 2.1.2 客户画像的常见方法
- 2.1.3 数据从哪儿来
- 2.1.4 应用场景

2.2 需求满足（30min）

- 2.2.1 主动性需求：客户想要什么
- 2.2.2 被动性需求：我们希望给客户什么

2.3 制造惊喜（30min）

- 2.3.1 大数据不止可以“杀熟”：抓住客户的潜在需求
- 2.3.2 “快了一点”与“快乐一点”：2B 与 2C

Part 2 大数据营销与数据化客户管理(CRM)

1 大数据营销的常见应用（30min）

- 1.1 高效获客
- 1.2 精准营销
- 1.3 流失预警
- 1.4 产品线管理
- 1.5 客户生命周期维护

2 高效获客（60min）

- 2.1 常用获客指标
- 2.2 低端客户过滤器：识别“羊毛党”的常用方法(含案例剖析)
- 2.3 CRM 系统与新客获取(含案例剖析)

3 精准营销 (60min)

- 3.1 常用营销指标
- 3.2 大数据活动方案设计与效果预测
- 3.3 CRM 客户标签完善路径与组合使用(含案例剖析)
- 3.4 利用 CRM 系统进行客户 A/B 试验(含案例剖析)

4 流失预警 (60min)

- 4.1 常用流失预警指标
- 4.2 客户流失潜伏期识别(含案例剖析)
- 4.3 CRM 系统与流失挽回的应用(含案例剖析)

5 产品线管理 (60min)

- 5.1 常用产品销售指标
- 5.2 客户覆盖率及产品线竞争力分析(含案例剖析)
- 5.3 结合 CRM 系统, 产品上下架业务影响预测(含案例剖析)

6 客户生命周期维护 (60min)

- 6.1 客户的两个生命周期
- 6.2 大数据在生命周期不同阶段的应用案例(含案例剖析)
- 6.3 CRM 系统与客户生命周期管理(含案例剖析)

7 CRM 系统的应用实例 (30min)

- 7.1 闭环 CRM 系统案例(含案例剖析)
- 7.2 “云系统”案例(含案例剖析)

Part 3 金融领域的数字化应用

一 客户篇 (3 h)

1. 整体了解: KYC 模型原理解析 (60 min)

- 1.1 主观因子

- 1.2 客观因子
 - 1.3 评分原理
 - 1.4 应用举例
 - 2. 新客：潜力筛选（60 min）
 - 2.1 渠道分析
 - 2.2 羊毛党分析
 - 2.3 客户价值评分与成本因子
 - 3. 老客：价值提升（60 min）
 - 3.1 数据化客户档案
 - 3.2 客户成长路径
 - 3.3 客户社交网络分析
- ## 二 产品篇（3 h）
- 1. 产品体系规划（60 min）
 - 1.1 产品分类
 - 1.2 产品网络分析
 - 1.3 引流产品、中途岛产品与粘性产品
 - 3. 组合推荐（60 min）
 - 3.1 常用推荐模型
 - 3.2 推荐效果评估
 - 3.3 结合客户标签的推荐
 - 3.4 结合行为数据的客户 A/B 试验
 - 3. 风险管理（60 min）
 - 3.1 异常行为识别
 - 3.2 潜在风险敞口管理
 - 3.3 平衡型风险管理

Part 4 应用方法与工具

由于时间限制，这部分以介绍每个工具的基本功能和应用场景为主，包括简单的操作展示和基础类的应用。后续建议根据实际需要重点选择学习某一工具。

一 分析类工具（3 h）

1. Excel (60 min)
 - 1.1 Excel 日常数据处理流程
 - 1.2 常用技巧分享
 - 1.3 案例：自动化数据处理模板设计
 2. R (60 min)
 - 2.1 R 的介绍：主要特点与功能
 - 2.2 常用处理流程和场景
 - 2.3 案例：R 语言做客户分类
 3. Python (60 min)
 - 3.1 Python 的介绍：主要特点与功能
 - 3.2 常用处理流程和场景
 - 3.3 案例：智能作诗
- 二 展示类工具 (3 h)
1. Excel (60 min)
 - 1.1 Excel 报表的设计思路
 - 1.2 案例：监控周报设计
 2. PPT (60 min)
 - 2.1 PPT 报表的设计思路
 - 2.2 案例：分析型月报设计
 3. Tableau (60 min)
 - 3.1 Tableau 介绍：主要特点与功能
 - 3.2 案例：交互式报表设计

八、大数据治理实战（4.5 天）

一、培训收益

通过此次课程培训，可使学习者获得如下收益：

1. 掌握数据治理、DAMA DMBOK 数据管理知识体系核心内容
2. 理解数据架构设计与数据标准、元数据管理等的基本知识；
3. 掌握企业级数据架构下的数据建模方法；
4. 理解数据质量、数据能力成熟度 DMM、主数据管理、数据安全等内容；
5. 理解数据治理落地实施最佳实践，各数据管理核心领域的重点、难点的深度解析和行业最佳实践案例；
7. 培养数据治理、数据管理的问题思考和方案设计、实施执行能力。

二、培训对象

学习者的职业定位：

1. CIO 首席信息官
2. CDO 首席数据官
3. 业务部门的信息化负责人与业务骨干
4. 数据管理团队、数据服务及数据运营团队
5. 数据资产管理产品经理及数据管理专家

三、培训特色

1. 理论与实践相结合、案例分析与理论穿插进行；
2. 专家精彩内容解析、学员专题讨论、分组研究；
3. 通过全面知识理解、专题技能掌握和安全实践增强的授课方式。

四、日程安排

第一天 上午 大数据治理目标与定位

1. 什么是数据治理？
2. 数据治理的基本理论
3. 数据管理的基本概念
4. 国内外数据治理发展及趋势

第一天 下午 数据架构管理

1. 企业架构基本理论
2. TOGAF 企业架构概述

3. 什么是数据架构?
4. 数据架构的基本概念及内容组成
5. 数据架构的建设方法
6. 数据架构与数据治理之间的关系
7. 数据架构的相关实践案例

第二天 上午 数据标准管理

1. 数据标准的基本概念
2. 数据标准的分类
3. 业务术语标准化
4. 参考数据和主数据标准化
5. 数据元标准化
6. 指标数据标准化
7. 数据标准的建设方法
8. 数据标准与数据治理之间的关系
9. 数据标准化的案例

第二天 下午 元数据管理

1. 元数据的基本概念
2. 元数据的分类
3. 元数据管理
4. 元数据管理工具
5. 元数据管理案例，构建企业数据资源目录

第三天 上午 数据质量管理

1. 数据质量基本概念
2. 全面数据质量管理及通用方法
3. 数据质量评估维度
4. 数据质量管理手段及工具
5. 数据质量管理案例

第三天 下午 数据能力成熟度 DMM

1. DMM 数据能力成熟度评估模型
2. DMM 数据能力成熟度评估方法及过程

3. DMM 数据能力成熟度评估案例

第四天 上午 主数据管理

1. 参考数据和主数据的基本概念
2. 主数据管理的主要内容和实施难点
3. 主数据管理工具
4. 主数据管理方法及实施方案
5. 主数据管理案例

第四天 下午 数据安全治理

1. 数据安全治理基本概念及内容框架
2. 数据安全分级
3. 数据安全策略
4. 数据安全治理
5. 数据安全审计
6. 数据安全治理案例

第五天 上午 行业数据治理实践

1. 能源行业数据治理实践分享
2. 银行、金融行业数据治理实践分享
3. 通信行业数据治理实践分享

九、如何撰写有价值的数据分析报告（1-2天）

篇	章	内容
Part 1 理论篇	数据报告的核心	<ul style="list-style-type: none"> • 目标 • 原则 • 效果
	数据报告的基本步骤	<ul style="list-style-type: none"> • 数据收集：内部数据与外部数据 • 数据处理：清洗与筛选 • 确定方法：分析与模型、实验与分组 • 总结结论：解释型、结论型与预测型 • 结果呈现：时效性、互动性与准确性
Part 2 思维篇	数据分析的本质： 解决问题	<ul style="list-style-type: none"> • 问题的三种形式 • 原因寻找型问题 • 风险防范型问题 • 追求理想型问题
	数据分析的路径： 从业务到抽象，再从抽象回归业务	<ul style="list-style-type: none"> • 战略规划：市场与竞对分析 • 客户管理：客户特征画像 • 产品设计：品牌价值主张与产品线规划 • 活动营销：精准营销与增长黑客 • 业务运营：客群维护与日常监控
Part 3 技巧篇	图表选择	<ul style="list-style-type: none"> • 单维度图表 • 双维度图表 • 多维度图表 • 综合信息图表
	常见问题	<ul style="list-style-type: none"> • 如何将问题抽象成数据的语言 • 如何选择恰当的分析维度 • 如何设计实验方案：对照组与实验组的选择 • 如何评估最终效果：常用指标设定与调整 • 缺失数据怎么处理：常用填补方法

		<ul style="list-style-type: none"> • 结果不符合预期怎么处理：寻因法与纠错法 • 陌生问题没有思路怎么处理：矩阵式分析与图形化思考 • 考虑的方面过多怎么处理：降维与独立变量法
Part 4 场景篇	方案介绍型报告	<ul style="list-style-type: none"> • 战略规划 • 活动营销 • 产品设计
	结果展示型报告	<ul style="list-style-type: none"> • 业绩展示 • 活动效果 • 市场分析
	原因剖析型报告	<ul style="list-style-type: none"> • 业绩问题 • 预警评估 • 落差归因
	常规周期型报告	<ul style="list-style-type: none"> • 汇报类 • 监控类 • 高管类
	创意分享型报告	<ul style="list-style-type: none"> • 团队内部 • 公司内部 • 市场客户
Part 5 工具篇	数据类报表：Excel	<ul style="list-style-type: none"> • 报表设计 • 常用公式 • 案例
	数据类报告：PPT	<ul style="list-style-type: none"> • 报告设计 • 结构与内容 • 案例
	交互式报告：Tableau	<ul style="list-style-type: none"> • 设计思路：数据来源、交互环节、逻辑关联 • 基本技巧 • 案例
Part 6	基于客户相关行业设计，	- 分析情景

综合案例篇	一般包括如下部分	<ul style="list-style-type: none">- 需求确定- 报告类型- 工具选择- 结构拆分- 分析过程（节选展示）- 结果呈现（节选展示）
-------	----------	---

