目录

[小白的数据分析之路 3](#_Toc511638920)

[课程特色： 3](#_Toc511638921)

[课程大纲： 4](#_Toc511638922)

[模块一：数据分析必备Excel工具 4](#_Toc511638923)

[第一部分 函数学习 4](#_Toc511638924)

[第1节 认识公式 4](#_Toc511638925)

[第2节 公式中的运算符和数据类型 4](#_Toc511638926)

[第3节 单元格引用类型 5](#_Toc511638927)

[第4节 跨工作表引用 5](#_Toc511638928)

[第5节 表格和结构化引用 5](#_Toc511638929)

[第6节 认识Excel函数 5](#_Toc511638930)

[第7节 函数的输入 5](#_Toc511638931)

[第8节 公式结果的检验、验证和限制 5](#_Toc511638932)

[第9节 文本处理技术 6](#_Toc511638933)

[第10节 信息提取与逻辑判断 6](#_Toc511638934)

[第11节 数学计算 6](#_Toc511638935)

[第12节 日期和时间计算 7](#_Toc511638936)

[第13节 查找与引用 7](#_Toc511638937)

[第14节 统计与求和 8](#_Toc511638938)

[第15节 数组公式 8](#_Toc511638939)

[第二部分：透视表与图表 9](#_Toc511638940)

[第三部分：Excel数据分析 9](#_Toc511638941)

[第四部分：综合案例 10](#_Toc511638942)

[模块二：数据提取必备SQL工具 10](#_Toc511638943)

[第一部分：SQL学习 10](#_Toc511638944)

[第1节：数据库和SQL 10](#_Toc511638945)

[第2节：查询基础 11](#_Toc511638946)

[第3节：聚合与排序 12](#_Toc511638947)

[第4节：数据更新 14](#_Toc511638948)

[第5节：复杂查询 14](#_Toc511638949)

[第6节：函数、谓词、 15](#_Toc511638950)

[第7节：集合运算 16](#_Toc511638951)

[第二部分：SQL综合应用案例 17](#_Toc511638952)

[模块三：数据建模工具python 17](#_Toc511638953)

[第一部分：python基础语法学习 17](#_Toc511638954)

[第1节：基础知识 18](#_Toc511638955)

[第2节：列表和元组 18](#_Toc511638956)

[第3节：使用字符串 19](#_Toc511638957)

[第4节：字典 20](#_Toc511638958)

[第5节：条件、循环及其他语句 20](#_Toc511638959)

[第6节：函数 22](#_Toc511638960)

[第7节：获取外部数据 22](#_Toc511638961)

[第二部分：python重点模块学习 24](#_Toc511638962)

[第三部分：python可视化 24](#_Toc511638963)

[第四部分：python综合应用案例 24](#_Toc511638964)

[模块四：数据分析之风控模型（结合python） 24](#_Toc511638965)

[第一部分：关于sklearn学习 25](#_Toc511638966)

[第二部分：逻辑回归评分卡模型的构建、优化与落地 25](#_Toc511638967)

[第三部分：cart、xgboost模型的构建、优化与实际应用 26](#_Toc511638968)

### 

### 小白的数据分析之路

#### 课程特色：

针对入门。课程体系设计的初衷是针对入门级学员，从最基础开始，循序进阶；我们始终认为，基础不牢，地动山摇，所以整体的课程设计及其强调数据分析基础的学习；

课程全面。涵盖基础知识、数据处理、可视化、数据建模，并且涉及模型建立后如何落地、落地后如何跟踪模型效果等，真正从实践层面考虑；

化繁为简。市面培训课程为广泛吸纳学员，课程设计范围非常宽泛，导致直接的结果是，学员学完后感觉什么都会，但真正实践中，却无从下手；本课程仅针对入门学员设计，摒弃眼花缭乱的数据分析工具以及所谓高大尚模型，而专精于实践中常用工具和建模方法，讲透讲彻底，便于学员一通百通；

一线数据分析实战老师。授课老师都是从事多年数据分析经验背景，并且有丰富培训团队经验，多次在公司为多部门以及集团层面开展数据培训，深谙入门学员成长的过程，课程设计符合入门学员循序成长的要求；

### 课程大纲：

### 模块一：数据分析必备Excel工具

毫不夸张的说，大部分人使用到的Excel功能不到其所有功能的10%；Excel作为一款优秀的办公软件，可以完成日常80%数据处理、分析甚至建模工作。对于任何一名数据分析人员，掌握Excel函数、透视表、图表是入门数据分析行业最基础的要求。

通过以下课程学习，让学员掌握如下能力：

熟练使用常用函数处理日常数据问题，如脏数据处理、异常处理、基础特征衍生、简单模型构建等；同时，熟练通过透视表汇总数据，并通过适合的图表展示分析结果。

#### 第一部分 函数学习

##### 第1节 认识公式

1.1 公式和函数的概念

1.2 公式的输入、编辑与删除

1.3 公式的复制与填充

##### 第2节 公式中的运算符和数据类型

2.1 认识运算符

2.2 数据类型的转换

##### 第3节 单元格引用类型

3.1 认识单元格引用

3.2 相对引用、绝对引用和混合引用

3.3 多单元格和单元格区域的引用

##### 第4节 跨工作表引用

4.1 引用其他工作表区域

##### 第5节 表格和结构化引用

5.1 创建表格

5.2 表格的计算

5.3 结构化引用

##### 第6节 认识Excel函数

6.1 函数的概念

##### 第7节 函数的输入

7.1 输入函数的几种方式

##### 第8节 公式结果的检验、验证和限制

8.1 使用公式的常见问题

8.2 公式结果的检验和验证

##### 第9节 文本处理技术

9.1 接触文本数据

9.2 文本的合并

9.3 文本值的比较

9.4 大小写、全角半角转换

9.5 字符与编码转换

9.6 字符串提取

9.7 查找字符

9.8 替换字符或字符串

9.9 格式化文本

##### 第10节 信息提取与逻辑判断

10.1 单元格信息函数

10.2 常用IS类判断函数

10.3 逻辑判断函数

10.4 屏蔽函数公式返回的错误值

##### 第11节 数学计算

11.1 取余函数

11.2 数值取舍函数

11.3 四舍五入函数

11.4 随机函数

##### 第12节 日期和时间计算

12.1 认识日期和时间数据

12.2 日期函数

12.3 认识DATEDIF函数

12.4 返回日期值的中文短日期

12.5 星期相关函数

12.6 工作日相关函数

12.7 时间的计算

12.8 时间和日期函数的综合运用

##### 第13节 查找与引用

13.1 行号和列号函数

13.2 基本的查找函数

13.3 特殊的查找函数

13.4 常用的定位函数

13.5 认识OFFSET函数

13.6 理解INDEX函数

13.7 了解CHOOSE函数

13.8 转置数据区域

13.9 使用公式创建超链接

13.10 提取公式字符串

13.11 获得单元格地址

13.12 查找引用函数的综合应用

##### 第14节 统计与求和

14.1 认识COUNT函数

14.2 认识COUNTA函数

14.3 认识COUNTBLANK函数

14.4 条件计数函数

14.6 条件求和函数

14.7 均值函数和条件均值函数

14.8 频率统计函数

14.9 极值与中值函数

14.10 其他常用统计函数

##### 第15节 数组公式

15.1 理解数组

15.2 数组公式与数组运算

15.3 数组构建及填充

15.4 条件统计应用

15.5 数据提取技术

15.6 数据筛选技术

15.7 利用数组公式排序

#### 第二部分：透视表与图表

1.基本图表介绍与制作

2.数据透视表

3.切片器

4.条件格式

5.数据透视表与切片器的链接

5.单元格区域、图表与控件间的链接

6.数据透视表、图表与控件间的链接

7.专业制图：突破Excel基础制图功能限制

#### 第三部分：Excel数据分析

1. 数据分析流程介绍
2. 对比分析
3. 分组分析
4. 结构化分析
5. 平均分析
6. 综合分析
7. 杜邦分析
8. 漏斗分析
9. 矩阵分析
10. Excel高级数据分析

#### 第四部分：综合案例

### 模块二：数据提取必备SQL工具

数据分析的第一步是获取数据，课程设计之初，我们摒弃了市面上热火朝天的爬虫内容，而专精于获取企业已有数据的技能；主要原因是，对于数据分析人员而言，爬虫并非其核心能力要求，甚至并非数据分析的工作；与其千辛万苦获取外部效果甚微的数据，不如切实用好自有数据，而SQL几乎是每个公司对数据分析人员的基本要求。

#### 第一部分：SQL学习

##### 第1节：数据库和SQL

数据库是什么

我们身边的数据库

为什么DBMS那么重要

DBMS的种类

数据库的结构

RDBMS的常见系统结构

表的结构

SQL概要

标准SQL

SQL语句及其种类

SQL的基本书写规则

##### 第2节：查询基础

SELECT语句基础

列的查询

查询出表中所有的列

为列设定别名

常数的查询

从结果中删除重复行

根据WHERE语句来选择记录

注释的书写方法

算术运算符和比较运算符

算术运算符

需要注意NULL

比较运算符

对字符串使用不等号时的注意事项

不能对NULL使用比较运算符

逻辑运算符

NOT运算符

AND运算符和OR运算符

通过括号强化处理

逻辑运算符和真值

含有NULL时的真值

##### 第3节：聚合与排序

对表进行聚合查询

聚合函数

计算表中数据的行数

计算NULL之外的数据的行数

计算合计值

计算平均值

计算最大值和最小值

使用聚合函数删除重复值（关键字DISTINCT）

对表进行分组

GROUPBY子句

聚合键中包含NULL的情况

使用WHERE子句时GROUPBY的执行结果

与聚合函数和GROUPBY子句有关的常见错误

为聚合结果指定条件

HAVING子句

HAVING子句的构成要素

相对于HAVING子句，更适合写在WHERE子句中的条件

对查询结果进行排序

ORDERBY子句

指定升序或降序

指定多个排序键

NULL的顺序

在排序键中使用显示用的别名

ORDERBY子句中可以使用的列

不要使用列编号

##### 第4节：数据更新

数据的插入（INSERT语句的使用方法）

什么是INSERT

INSERT语句的基本语法

列清单的省略

插入NULL

插入默认值

从其他表中复制数据

数据的删除（DELETE语句的使用方法）

DROPTABLE语句和DELETE语句

DELETE语句的基本语法

指定删除对象的DELETE语句（搜索型DELETE）

数据的更新（UPDATE语句的使用方法）

UPDATE语句的基本语法

指定条件的UPDATE语句（搜索型UPDATE）

使用NULL进行更新

多列更新

##### 第5节：复杂查询

子查询

子查询和视图

子查询的名称

标量子查询

标量子查询的书写位置

使用标量子查询时的注意事项

关联子查询

普通的子查询和关联子查询的区别

关联子查询也是用来对集合进行切分的

结合条件一定要写在子查询中

##### 第6节：函数、谓词、

各种各样的函数

函数的种类

算术函数

字符串函数

日期函数

转换函数

谓词

什么是谓词

LIKE谓词——字符串的部分一致查询

BETWEEN谓词——范围查询

ISNULL、NULL——判断是否为NULL

IN谓词——OR的简便用法

使用子查询作为IN谓词的参数

EXIST谓词

CASE表达式

什么是CASE表达式

CASE表达式的语法

CASE表达式的使用方法

##### 第7节：集合运算

表的加减法

什么是集合运算

表的加法——UNION

集合运算的注意事项

包含重复行的集合运算——ALL选项

选取表中公共部分——INTERSECT

记录的减法——EXCEPT

联结（以列为单位对表进行联结）

什么是联结

内联结——INNERJOIN

外联结——OUTERJOIN

3张以上的表的联结

交叉联结——CROSSJOIN

联结的特定语法和过时语法

#### 第二部分：SQL综合应用案例

### 模块三：数据建模工具python

Python无疑是目前机器学习甚至AI主流工具，虽然通过Excel VBA可以实现几乎所有建模方法，但其难度也可想而知。

通过对python的学习，掌握数据处理、特征衍生以及模型构建等技能，同时包括python可视化以及主流模块的学习。

#### 第一部分：python基础语法学习

##### 第1节：基础知识

交互式解释器

算法是什么

数和表达式

变量

语句

获取用户输入

函数

模块

注释

字符串

单引号字符串以及对引号转义

拼接字符串

字符串表示str和repr

长字符串、原始字符串和字节

##### 第2节：列表和元组

序列概述

通用的序列操作

索引

切片

序列相加

乘法

成员资格

列表：Python的主力

函数list

基本的列表操作

列表方法

元组：不可修改的序列

##### 第3节：使用字符串

字符串基本操作

设置字符串的格式

替换字段名

基本转换

宽度、精度和千位分隔符

符号、对齐和用0填充

字符串方法

center

find

join

lower

replace

split

strip

translate

判断字符串是否满足特定的条件

##### 第4节：字典

字典的用途

创建和使用字典

函数

基本的字典操作

将字符串格式设置功能用于字典

字典方法

##### 第5节：条件、循环及其他语句

再谈print和import

打印多个参数

导入时重命名

赋值魔法

序列解包

链式赋值

增强赋值

代码块：缩进

条件和条件语句

这正是布尔值的用武之地

有条件地执行和if语句

else

elif

代码块嵌套

更复杂的条件

循环

while

for

迭代字典

一些迭代工具

跳出循环

循环中的else

简单推导

使用del

使用exec

##### 第6节：函数

自定义函数

给函数编写文档

参数魔法

值从哪里来

我能修改参数吗

关键字参数和默认值

收集参数

分配参数

作用域

递归

##### 第7节：获取外部数据

读写 CSV 文件

基本字符串分析是如何失败的

筛选特定的行

行中的值满足某个条件

行中的值属于某个集合

选取特定的列

列索引值

列标题

选取连续的行

添加标题行

读取多个 CSV 文件

从多个文件中连接数据

读写 Excel 文件

筛选特定行

选取特定列

读取工作簿中的所有工作表

在所有工作表中筛选特定行

在所有工作表中选取特定列

在 Excel 工作簿中读取一组工作表

处理多个工作簿

工作表计数以及每个工作表中的行列计数

连接mysql数据库

#### 第二部分：python重点模块学习

pandas模块详解（单独深入学习）

数据筛选

多条件索引

缺失处理

分组处理数据

数据合并与重构

进一步探讨pandas的深入学习

#### 第三部分：python可视化

#### 第四部分：python综合应用案例

### 模块四：数据分析之风控模型（结合python）

在风险控制领域，模型的应用非常广泛且成熟，通过这个模块，学员能够掌握常用模型的原理及使用场景，并且了解模型后期的落地与效果跟踪，在掌握一个或者几个模型后，再拓展学习其他模型，将是很容易的事情；但相反，如果开始就贪多，仅了解模型的大概便浅尝辄止，既不利于在实践中学以致用，也不利于拓展学习。

#### 第一部分：关于sklearn学习

#### 第二部分：逻辑回归评分卡模型的构建、优化与落地

1. 关于评分卡模型
   1. 三大风险介绍（市场风险、操作性风险、信用风险）
   2. 评分卡模型简介
   3. ABC卡
2. 特征衍生
   1. 训练集与测试集
   2. 特征构造
   3. 分箱和WOE转换

1. 入模特征筛选
   1. IV
   2. 变量分析
   3. 相关性分析
   4. 多重共线性

1. 逻辑回归模型
   1. 逻辑回归介绍
   2. 逻辑回归评分卡模型
   3. 分数

1. 模型优化与落地
   1. 常用评价指标ks auc psi
   2. 如何选择分值
   3. 模型迭代
2. 更多关于风控
   1. 风控系统流程介绍
   2. 关于准入与反欺诈环节
   3. 关于成本核算

#### 第三部分：cart、xgboost模型的构建、优化与实际应用

1. cart树介绍及构建信用模型
2. xgboost介绍及构建信用模型
3. 不同模型效果的对比