

## 第十二章

### 题目一 简单编程题（创建DataFrame数组并进行相应操作）

---

创建一个30\*6的DataFrame数组，元素由70~100之间均匀分布的随机整数构成，行标签按030201（初三.二班1号）~030230格式顺序排列，列标签分别为语文、数学、英语、物理、化学、计算机。

- 1、输出其纯数据、行标签、列标签、形状、大小和数据类型；
- 2、获取全班数学成绩、获取学号为030205的同学的所有成绩；
- 3、增加总成绩的新列，并建立按总成绩降序排列的副本（注意是获得副本，不是获得视图），切片获得前十名学生的全部成绩；
- 4、创建一个DataFrame对象（记为B），行标签与上文DataFrame对象（记为A）一致，列标签为性别，数据为30个学生的随机性别，将A和B进行水平合并，获得新的DataFrame对象（记为C）；
- 5、输出数据C的info和describe信息，尝试自定义my\_describe，输出自己感兴趣的统计信息；
- 6、按性别进行分组，对比男生女生所有科目及总成绩的平均值。

### 题目二 复杂编程题（DataFrame数组操作）

---

下载titanic数据集，执行下列操作：

```
import seaborn as sns
import sys
titanic = sns.load_dataset("titanic")
titanic.head()
```

- 1、获得一个删除了无年龄数据的所有行的副本；
- 2、创建一个名为Age的Series对象，其数据来源于对数据集中的年龄按下列规则进行映射（参照行星数据集案例中decade的处理办法）：

```
If <10 : "0s"
```

```
elif <20 : "10s"
```

```
...
```

```
elif <60 : "50s"
```

```
...
```

3、通过sex和Age对titanic数据集进行分组，获得不同性别、不同年龄段乘客的幸存比例，请分别使用groupby和pivot\_table（如果直接用Age不行的话，换个思路）两种方法。

## 题目一 简单绘图题 (Matplotlib)

---

绘制以下两个函数的图像

1、设定一个你喜欢的绘图风格；

$$y = \frac{1}{1+e^{-x}}$$
$$y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

2、绘制折线图：

- 对线条的颜色、形状、粗细进行设置；
- 对数据点的形状、大小和颜色进行设置；
- 调整坐标轴的范围，调整坐标轴刻度的大小；
- 增加x轴、y轴的标签以及图像标题；
- 增加图例、添加箭头和文字；
- 将最终绘制好的图像保存到本地文件夹。

3、绘制散点图：

4、绘制柱形图：

- 利用函数1，生成一系列x, y，并绘制柱形图；
- 利用函数2，生成一系列x, y，并绘制横向柱形图；
- 利用函数1、2，生成一系列x, y，并绘制累加柱形图和并列柱形图；

5、多子图：

- 在函数1、2的基础上，增加正弦函数和余弦函数，绘制2\*2的多子图。

6、直方图：

- 构造一个标准正态分布，绘制频次直方图、概率密度直方图、累计概率直方图。

## 7、误差图：

- 利用函数 1，生成一系列x, y，随机生成一系列误差dy，绘制误差图。

## 题目二 简单绘图题（其他）

### 1、面向对象绘图

- 利用函数 1 和函数2，绘制一个画中画。

### 2、三维图形绘制

- 绘制一个薄圆柱体。

### 3、Seaborn 绘图

- 用 seaborn 风格绘制题目一中的部分图形，与matplotlib 风格进行对比。

### 4、Pandas 绘图

- 自己构建一个DataFrame 对象，直接绘制线形图、柱形图、散点图、直方图、多子图等。

```
iris = sns.load_dataset('iris')
```

## 第十四章

### 题目一、约会网站配对效果数据集的处理

韩梅梅想在婚恋网站上找一个合适的男朋友，婚恋网站经常给她进行推荐，经过一段时间，她把网站推荐给他的人选进行了分类，分别是：不喜欢的人、魅力一般的人和极具魅力的人。

其中，网站推荐人选时附带三个重要的特征，分别是：该男士每年的飞行距离、玩视频游戏所耗时间的百分比、每周消费冰激凌的公升数。

数据集存储在文件 `datingTestSet.txt` 中。请执行下列操作：

- 1、 读取数据，并将数据转换成 `DataFrame` 格式；
- 2、 查看数据集；
- 3、 根据是否需要，对数据集进行预处理，如：
  - 标签编码、处理缺失值、数据标准化；
- 4、 构建训练集和测试集
- 5、 使用 8 种不同算法，分别对数据集进行训练，获得分类模型，并用测试集进行测试，最后将预测结果存储到本地文件中。
- 6、 找一个表现较好的算法，对比舍弃一个不重要特征与否对模型性能的影响。
- 7、 找一个表现较好的算法，舍弃一个不重要特征后，将训练模型可视化。