**关于学习任务分配：**在大类上可从客户、课程和私教这三个方面来考虑，具体来说，可有以下分析角度：

1、不同性别客户的课程类型，健身热衷程度，健身目的等的差异，以定位不同性别的客户的需求。

2、不同年龄段客户的健身数量、健身类别、可支配时间参与健身活动等的差异，以定位不同年龄段的客户需求。

3、不同职业客户健身的热衷程度、健身类别的差异，以定位不同职业的客户的客户需求。

4、不同收入客户健身的热衷程度、健身类别等的差异，为不同收入的客户提供性价比 最合适的课程。

5、分析潜在会员的相关信息，以便将潜在客户转变成正式会员。

6、分析大客户性质（公司福利）的人群的健身偏好、热衷程度、健身类别等，以更好地留住大客户。

7、不同课程类别的受欢迎程度、销量高低等，研究课程类别和内容对客户的影响。

8、同一门课程不同教练的受欢迎程度，研究最大化发挥私教资源以提高健身所利润。

9、分析新客户与老客户之间的比例，进而分析该会所的管理服务问题。

10、不同地段的分公司的经营业绩，研究地域与销售盈利关系。

以上这些列出的分析角度我们在后期会重新选择，选取其中一部分进行深入研究。

上述十点分析角度，三个人分配，4, 3 3，根据这些分析角度，自己认为需要对数据如何处理，可以运用什么算法，可以列出一份文档，也可以开始简单分析

算法可从《可行性分析》中列出的算法去选择

# 《统计学——基本概念和方法》

统计学家做的许多工作都是关注一个变量是否影响另一个变量，把这种关注高阔为四个问题：

（1） 在数据中，变量之间是否有关系

（2） 变量之间的关系有多强

（3） 总体中是否有关系

（4） 观测到的关系是一种因果关系吗

## 1、随机性和规律性

统计工作的三个部分：

数据收集、数据分析和由数据做推断

统计学可以被定义为一组由（1）收集数据（2）分析数据（3）由数据得出结论而组成的概念、原则和方法

随机性和规律性是统计的两个重要概念

数据的收集分为两个步骤

第一步选择用于度量待考察的元素

第二步是实际的收集数据

## 2、数据的收集

观测值=真值+非统计错误+随机性

好数据是指根据合理、正确的统计原理收集到的数据

## 3、数据的描述：图和表

数据的可视化

图优性：

是指图能够在最短的时间内：用最少的笔墨，在最小的空间内，给观众最多的思想

## 4、数据的描述：计算汇总统计量

集中趋势的衡量（均值、中位数、众数）

差异的衡量（标准差和方差）

汇总值有一个主要的有点和缺点：

优点：汇总值会使数据高度的简单化

缺点：任何的简单化都意味着某些数据的丢失

标准差：重要的偏差，是到均值的一种平均距离

标准差&标准误差

由原始观察值算出的叫做标准差，由一组均值算出的叫做标准误差

标准误差：是很多不同样本的均值的标准差

## 5、概率

四个主要理论统计量：z t X2 F

P值意味着什么和人们怎样基于事件发生的概率来对数据做出决策这样问题为假设检验提供了舞台。

二项分布：只两种结果

Poisson分布：小概率事件，例如无安打比赛

超几何分布：当样本很少时，能用与分析两个分类变量

Z分布：标准正太分布，均值为0，标准差为1，钟型，中点两边各有50%的观察值，曲线下面有95%的面积在-1.96到1.96之间

T分布：与z分布曲线基本一致，但是正太分布的中部较高，T分布在水平轴上的收敛不像正太分布那么快。这个区别表明T分布在其均值周围的聚集程度要比正太分布要差一些

T分布的自由度越大，T分布曲线越接近正太分布，在自由度50时这两条曲线就几乎相同

X2分布：分布是偏斜的，非对称的，非负的

F分布：也是非负的，F取值大部分在0-5之间变化

注意：t X2 F变量都是从z变量中衍生出来的，所以使用这三种变量时，都已经事先假定了数据服从正太分布

## 6、做出结论：估计

辨明了样本统计量与总体参数之间的差别和从样本统计量中估计参数的方法。

点估计——》无偏估计&有偏估计

区间估计：是用于参数估计值的一个范围，一个区间比一个单值能提供更多的信息

## 7、做出结论：假设检验

零假设&备择假设

## 8、变量间的关系

对于问题一：考量样本数据的模式，如果发现某种关系，则提出问题二

对于问题二：计算变量之间关系的强度

对于问题三：建立一个变量之间没有关系的零假设并检验这个假设看是否拒绝它

问题四：两个变量之间

## 9、两个分类变量的X2分析

## 10、两个数值型变量的回归分析和相关分析

散点图表明变量之间是正相关还是负相关，相关洗漱室衡量关系的强度

## 11、Anova：一个分类变量和一个数量变量的方差分析

## 12、两个顺序变量的秩方法

## 13、多元分析

## 14、日常生活中的统计