

清洗和整理数据

分析数据的第一步应该是清洗和整理数据，也顺便在这个过程中熟悉一下数据的基本情况。通过探索，我返现数据中存在一些明显的异常值，我对这些异常值进行了清洗和整理，大致情况如下：

1. 我用Today()函数结合问卷数据中的birthdate计算了每个用户的Age，发现最小值1岁，最大值119岁，我认为这是异常值，删掉了它们，共41条；
2. 关于how many hours of sleep per night，65和85是明显的异常值，因为一天才24小时，也做了删除处理；
3. 关于how many hours spend sitting per day，超过24小时的也做了删除处理；
4. 关于hours of learning/applying per week这两个字段，对于文字描述做了清理，其他模糊不清的描述，比如10-20，30+之内的，将其简单处理成纯数字，10+的就按10小时算，方便后面的分析；

要探索的问题

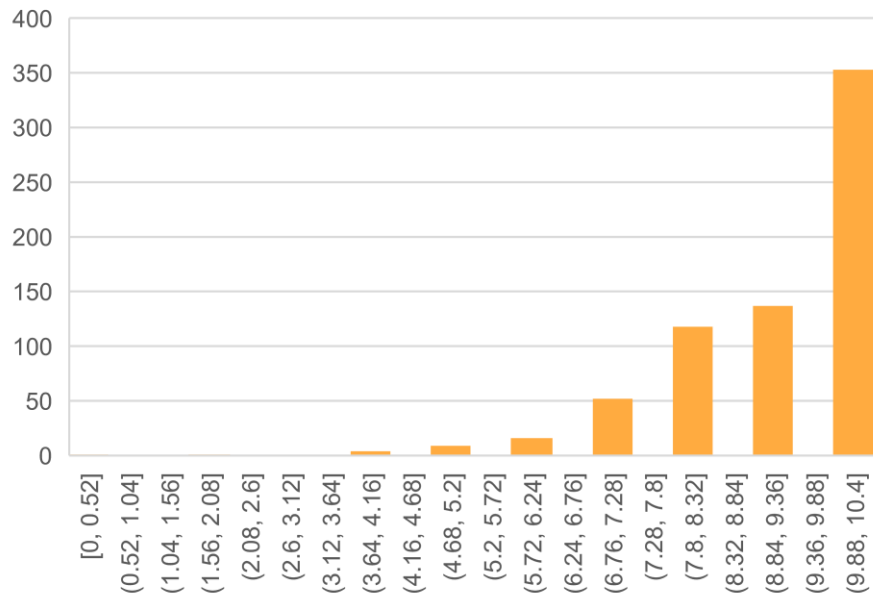
1. 学员们对Udacity的评价是怎样的？
2. 最受欢迎的纳米学位课程有哪些？
3. 不同职业学员选修纳米学位的情况是怎样的？
4. 不同学历的学员每周用于学习知识的时间有什么不同？

学员们对Udacity的评价是怎样的？

	Point
Minimum	0
Q1	8
Q2	10
Q3	10
Maximum	10

从直方图和五数总括上可以看出，学员们普遍对Udacity的纳米学位课程评价很高，75%的学员评分在8分以上，表示非常愿意向自己的亲朋好友推荐Udacity。

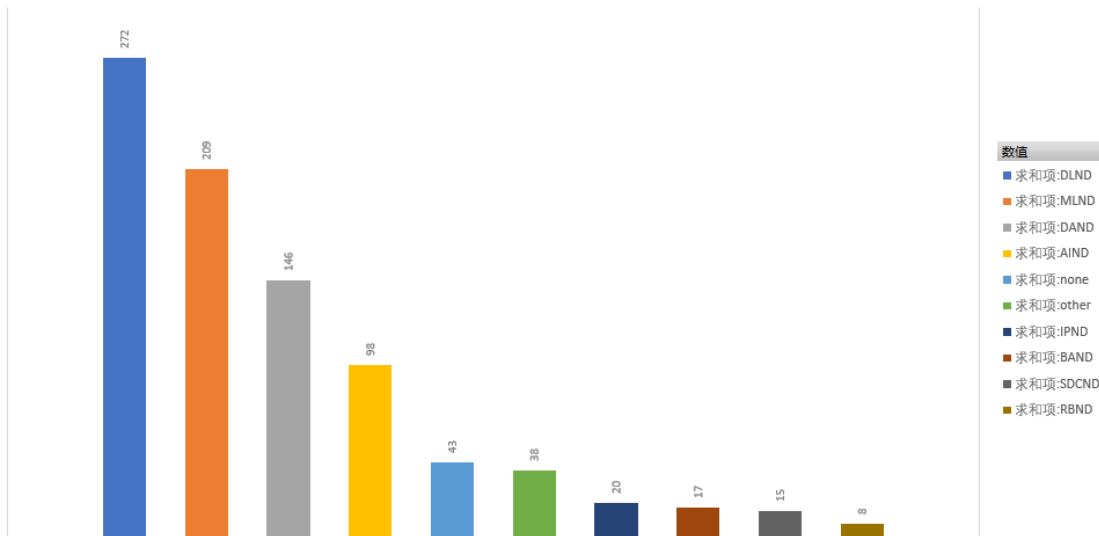
学员们对Udacity的评价



最受欢迎的纳米学位课程有哪些？

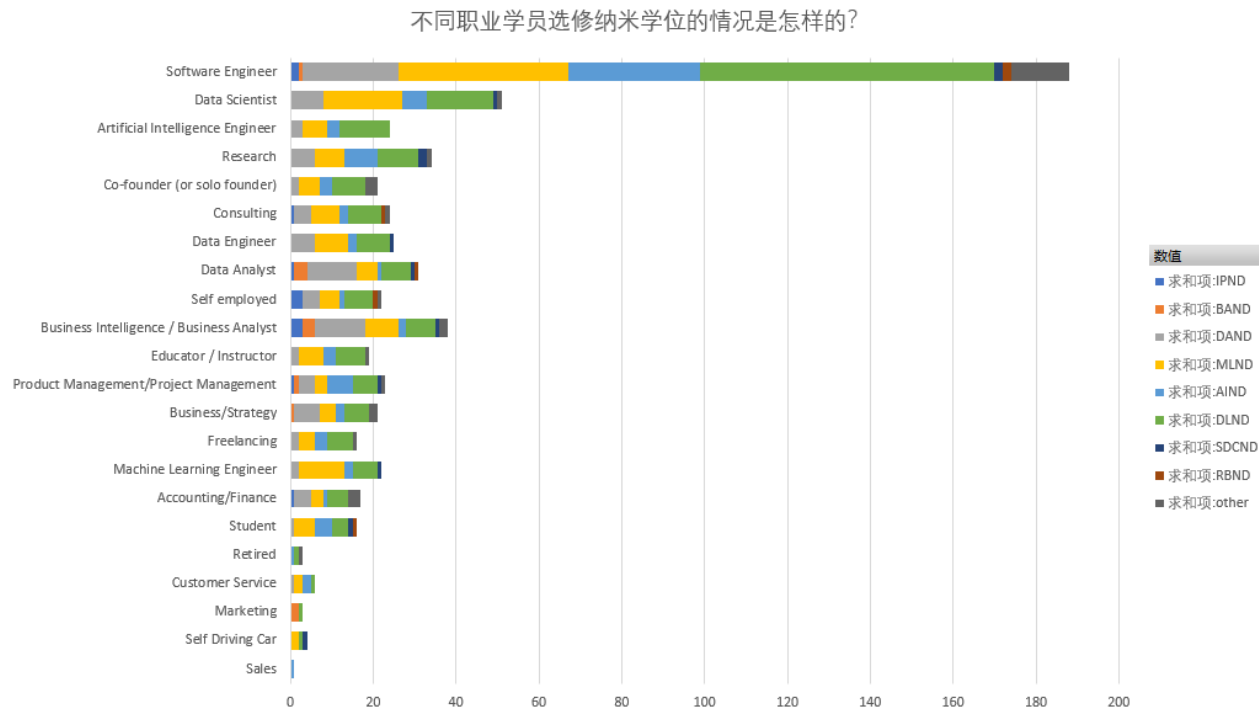
从图中可以看到，最受欢迎的是深度学习纳米学位，其次是机器学习，然后是数据分析，排第4名的是人工智能纳米学位。

最受欢迎的纳米学位课程有哪些？



不同职业学员选修纳米学位的情况是怎样的？

软件工程师对纳米学位的需求量最大，且这个岗位学习的最多的是机器学习和深度学习相关的纳米学位，软件工程师们对新技术抱有很大的热情，而职位是数据科学家的学员，也通过Udacity来学习机器学习和深度学习，职位是销售的学员选修最少。



不同学历的学员每周用于学习知识的时间有什么不同？

本科学历和硕士学历的学员投入的学习时间最多，远远超过其他学历的学员，而排第三的是博士学位的学员，学历为Associates（美国大学修满二年课程的肄业证书）的学员投入的时间最少。

不同学历的学员每周用于学习知识的时间有什么不同

