

问题一：参与问卷调查人群的学习动机有哪些？

学习动机情况

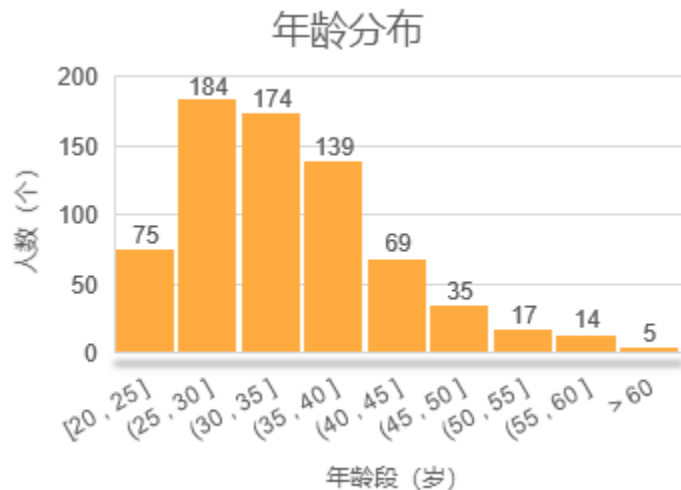
学习动机	人数
Start a new career in this field	415
Grow skills for my current role	312
Help move from academia to industry	88
Help prepare for an advanced degree	124
General interest in the topic (personal growth and enrichment)	411
Others	7

对数据集中的前6列数据进行了数据统计，前六列数据表示了参与问卷人群的学习动机，其中想要开始在这个领域中开始新的职业生涯的有415人，出于个人兴趣参与的有411人，想要在目前的状态下提升技能水平的有312人，想为更高的学位做准备有124个人，人想要从学术转为实践有88个，还有7个人出于其他原因。

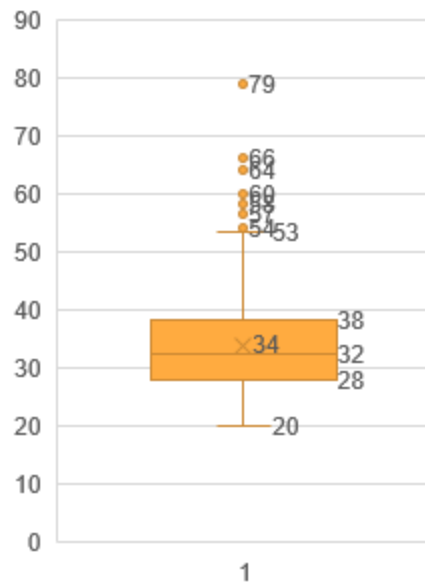
问题二：参与调查问卷人群的年龄分布情况如何？

各年龄段人数

年龄段	人数
20-25	75
25-30	184
30-35	174
35-40	139
40-45	69
45-50	35
50-55	17
55-60	14
>60	5



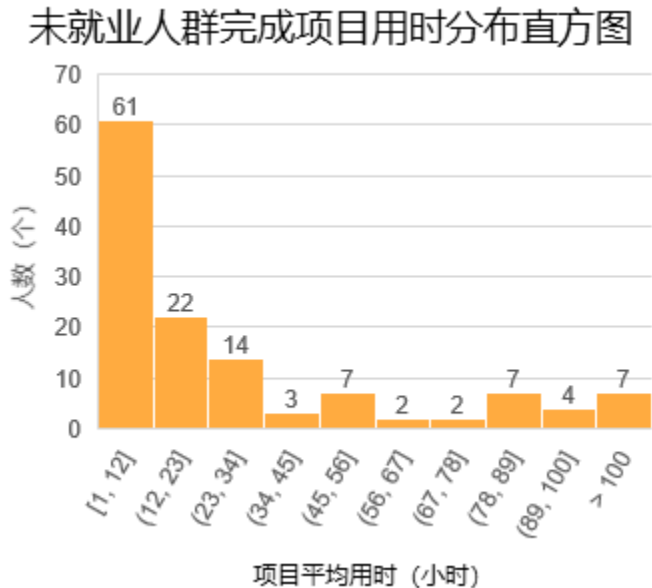
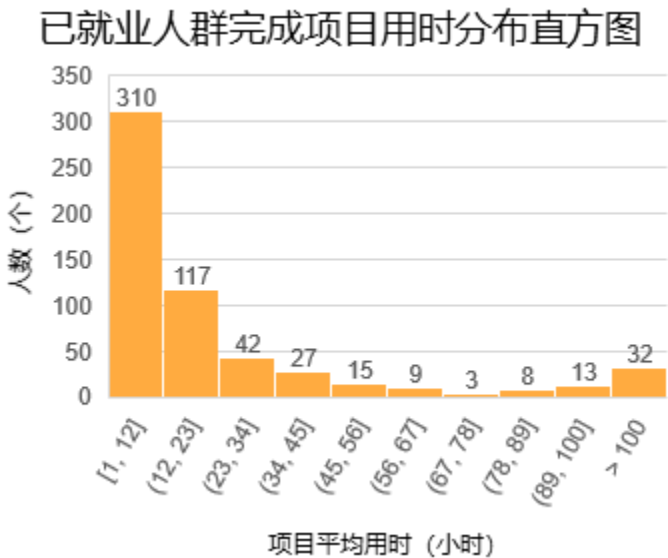
年龄箱形图



对数据集中的出生日期进行了数据清理，将其转换为对应年龄，并将数据集中出生日期为空以及很小的异常值剔除，得出的年龄分布统计数据在左表，从表中可以看出年龄段的众数是25-30这个年龄段。分布直方图如中图，整体呈右偏分布，箱型图如右图，从箱型图可以看出平均年龄为34，年龄中位数为32，大部分数据集中于28-38之间。

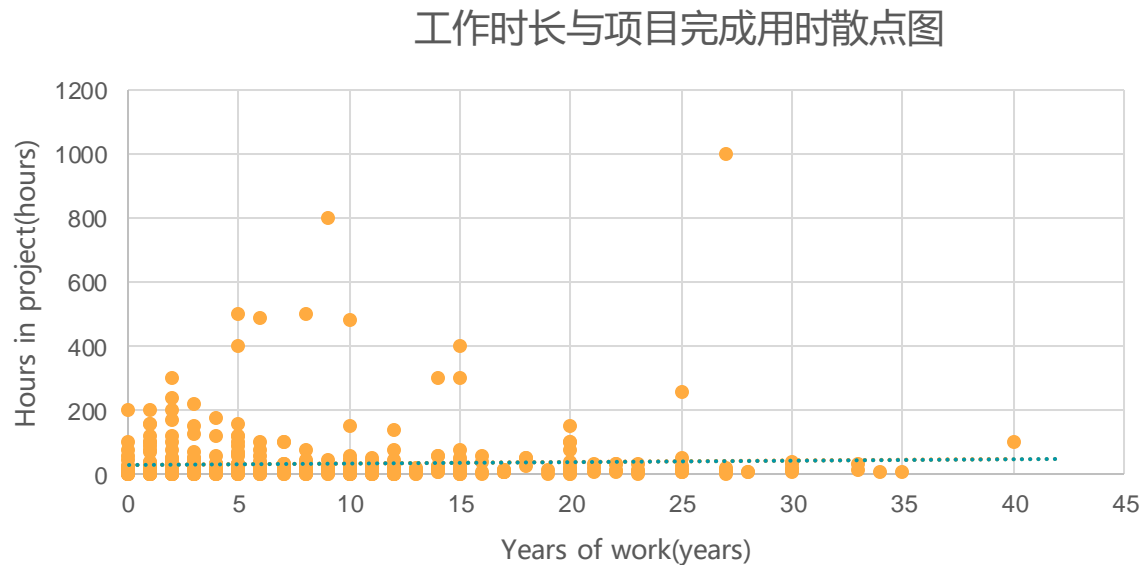
问题三：已就业的人群完成项目用时比未就业人群的多吗？

统计指标	已就业	未就业
平均值	32.88889	37.09302
标准误差	3.28615	6.286824
中位数	12	15
众数	10	10
标准差	78.8676	71.4046
方差	6220.099	5098.616
峰度	63.40554	28.12259
偏度	6.982068	4.929628
区域	999	499
最小值	1	1
最大值	1000	500
求和	18944	4785
观测数	576	129



对就业人群 (Are you employed = 1) 和非就业人群 (Are you employed = 0) 的数据进行清洗，将对应项目时间为空的数据删除，使用Excel的数据分析工具作出描述统计，可以看出已就业人群的项目完成用时平均值是32.89，标准差是78.87，众数是10，值域为999，未就业人群的项目完成用时平均值是37.09，标准差是71.4，众数是10，值域是499，已就业人群的项目完成时间要比未就业人群的要少，数据量多于未就业人群，离散程度更大，分别作出他们的分布直方图，都呈右偏分布。

问题四：工作年限与完成项目用时有什么样的关系？



筛选出所有已就业的人群，针对他们的工作年限（years of experience in the field of work）和完成项目的时间，使用各列的平均值填充空白值，绘制散点图如上图，趋势线如绿色的虚线描述，是一个大致正相关的关系。