本次实验的数据是旅行者评价数据集，每行为一条数据，不同列之间使用"|"符号分隔，从前到后，每列的含义及数据类型为：

1. string review_id
2. double longitude
3. double latitude
4. double altitude
5. string review_date
6. string temperature
7. double rating
8. string user_id
9. string user_birthday
10. string user_nationality
11. category user_career
12. double user_income

具体的，我们对本次实验内容所使用的属性如下：

1. 3.1节中的分层抽样所使用的属性A为user_career属性；
2. 3.2节中的奇异值过滤所使用的属性B为两个属性，分别是longitude和latitude。为了避免同学们使用额外的流程计算奇异值边界，给定longitude的有效范围为[8.1461259, 11.1993265]，latitude的有效范围为[56.5824856, 57.750511]，可以在代码中直接使用；
3. 3.3中，数据格式属性涉及到属性user_birthday和review_date，这些日期字段可能使用2018-03-21、2018/03/21、March 21, 2019这些格式，转换为哪种格式取决于同学们自己；temperature有华氏与摄氏两种，同样的，目标格式取决于同学们；需要归一化的属性则是rating；
4. 3.4中存在缺失值的属性为rating和user_income，根据先验知识，rating近似依赖于user_income、longitude、latitude和altitude，user_income近似依赖于user_nationality和user_career。对rating和user_income的填充可以利用这些依赖关系；

总结一下
1.分层抽样：user_career(10)
2.奇异值过滤：longitude(1)和latitude(2)，边界给出了
3.数据格式属性：user_birthday(8)和review_date(4)；temperature(5)
4.归一化：rating(6)
5.缺失值属性：rating(6)和user_income(11),在文件中以？占位
- rating依赖于user_income(11)、longtitude(1)、latitude(2)和altitude(3)
- user_income依赖于user_nationality(9)和user_career(10)