[EEG3235A]Python和数据科学

**Assignment 1**

[**作业以word文件发送我邮箱stu\_pythonds@126.com**](mailto:作业以word文件发送我邮箱stu_pythonds@126.com)**。**

**文件命名为：作业1-学号-姓名**

**截止日期：2019年9月17日8：00am**

下载和安装Python 3.7 Version（https://www.anaconda.com/distribution/）；下载Adult Data Set（adult.data, https://archive.ics.uci.edu/ml/datasets/Adult）；学习和运行lecture\_1.py。

1. 当数据中存在缺失值（Missing Value）时，有哪些处理方法？（查阅文献，给出参考文献）
2. 编写Python代码，统计Adult Data Set中的缺失值情况。比如，一共有多少个缺失值？有哪些特征('age', 'wc', 'fnlwgt', 'edu', 'edu-num', 'ms', 'occ', 'rs', 'race', 'sex', 'cap-g', 'cap-l', 'hpw', 'native'等)存在缺失值？有多少个样本存在缺失值？有多少个样本只有一个特征存在缺失值？有多少个样本有两个特征同时存在缺失值？（提示：可能会用到data.count()和data.wc.isnull()等函数及集合的一些操作）
3. 阅读关于《How to lie with statistics》的文章，谈谈你的看法。