**Daily Report**

10th March, 2018

1. **Cut**

对于一组离散化的数字，可以通过cut来进行区分，将其分归到各个组中,如：

In [154]: bins = [18, 25, 35, 60, 100]

In [155]: cats = pd.cut(ages, bins)

对于各个区间，与高数内容相同，对于小括号，代表开区间，对于中括号，则代表闭区间，对于开闭区间的选择，可以使用right=True/False进行

当只传入分块数量而没有传入分块的限界时，cut将根据数值的最大值、最小值和分块数量，对其进行平均分配

1. **Numpy.random.seed（）**

seed( ) 用于指定随机数生成时所用算法开始的整数值，如果使用相同的seed( )值，则每次生成的随即数都相同，如果不设置这个值，则系统根据时间来自己选择这个值，此时每次生成的随机数因时间差异而不同

1. **Permutation**

对于数组的随机排列，可以使用numpy.random.permutation()，该函数能够将数组的行随机调换顺序

通过ix-或者take（）功能，可以获取调换顺序后的数组

1. **Get\_dummies（）**

对于一个Dataframe，可以通过get\_dummies（）返回一个包含0或1的Dataframe，存在值处显示1，否则为0

1. **Findall，search，match**

Findall将返回所有匹配值；search将返回第一个匹配值，并且可知其开始与结束为止；match只有在开头处能匹配时才可使用

1. **Sub**

Sub可以用于替换；当一般表达式成功匹配时，将用输入的string替换匹配的内容