

## Quantum One 数据科学家笔试题目

附件Excel中是一张表名为LoanData的数据表及相关的数据字典DataDictionary， 请利用这张表完成以下问题：

1. 请以借款金额 (loan\_amt) 为目标变量建立一个线性模型。

\* 注：请将Python/R/SAS程序和结果单独列出，并简单解释模型结果

结论：在线性模型中，如下变量（按重要程度）可以用来Model loan\_amnt: 1. installment  
2. term 3.sub\_grade 4. mort\_acc 5. verification\_status 6. revol\_bal 7.purpose 8.annual\_inc  
9.acc\_open\_past\_24mths

**Possible Improvements:** 1.更好的处理missing value,具体方法前面code里面有提 2.根据需  
要选择不同的Statistical Indicator来筛选dependent variables(我用的是默认的sbc, Bayesian  
Information Criterion) 3. 分析各个dependent variable之间的Collinearity, 从而进一步筛选

2. 以借款利息 (int\_rate) 高低区分不同信用等级的借款人，以此为目标变量建立逻辑回归模型。

\* 注：请将Python/R/SAS程序和结果单独列出，并简单解释模型结果

结论：在线性模型中，如下变量（按重要程度）可以用来Model loan\_amnt: 1. installment  
2. term 3.sub\_grade 4. mort\_acc 5. verification\_status 6. revol\_bal 7.purpose 8.annual\_inc  
9.acc\_open\_past\_24mths

**Possible Improvements:** 1.更好的处理missing value,具体方法前面code里面有提 2.根据需  
要选择不同的Statistical Indicator来筛选dependent variables(我用的是默认的sbc, Bayesian  
Information Criterion) 3. 分析各个dependent variable之间的Collinearity, 从而进一步筛选

3. 以借款利息 (int\_rate) 高低区分不同信用等级的借款人，以此为目标变量建立一至俩个机器学习模型（GBM、Random Forrest、Neural Network、SVM等）。

\* 注：请将Python/R/SAS程序和结果单独列出，并简单解释模型结果