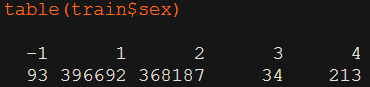
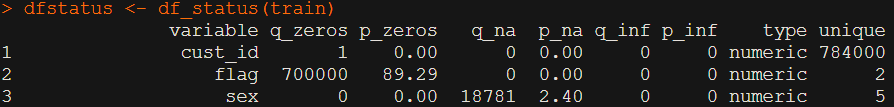
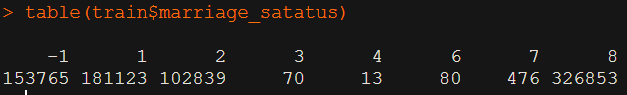
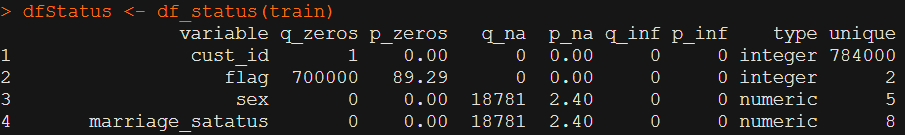
训练集里的feature包括了339个变量，其中cust\_id是定义客户的唯一id，flag字段为是否购买产品的label，其中第3-7个变量为客户基础属性， 第8-14个变量为账户数量， 第15-35个变量为账户交易情况，第36-48个变量为资产情况，第49-58个变量为贷款情况，第59-166及315-330个变量为渠道交易，第167-178个变量为渠道行为，第179-282个变量为第三方交易，第283-314个变量为自助设备交易，第331-338个变量为其他标识，最后一个变量为购买时间。

**训练集**

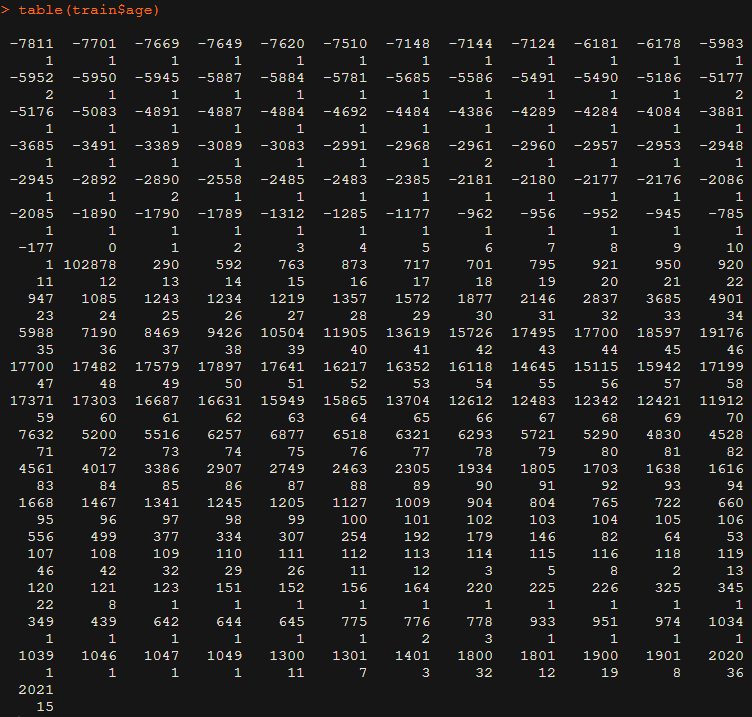
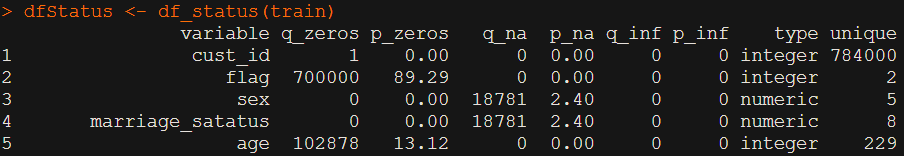
* Sex性别 （处理取值-1、3、4、NA）

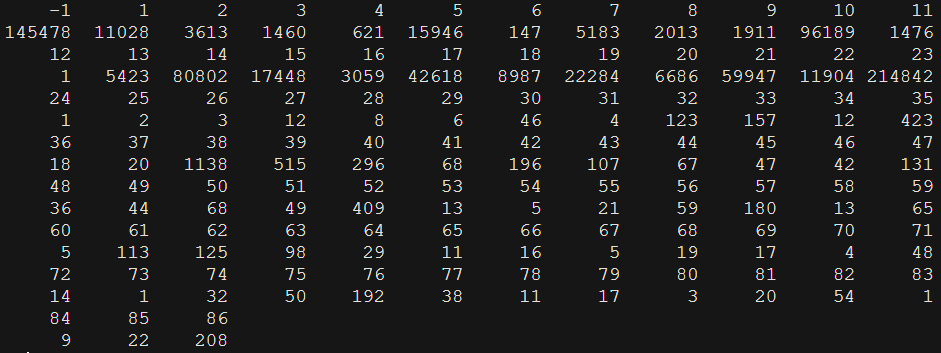
*  marriage\_satatus 结婚状态 （处理取值）

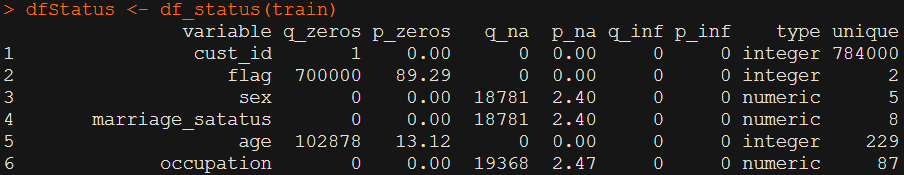
 

* Age年龄 （处理异常值）

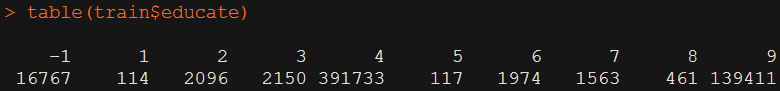
 

* Occupation 职业

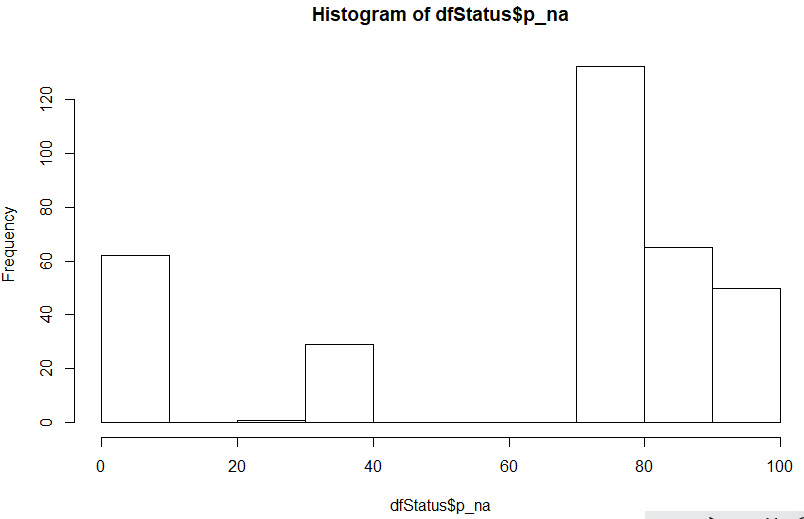




* Educate 教育

* 全部列缺失值比例分布



**测试集**