**《实用商业数据分析》项目描述**

**一、基本情况**

巴西银行(Banco do Brasil)，是巴西最大的银行，也是拉丁美洲最大的银行。它成立于1808年，总行设在巴西利亚，是巴西国有银行。巴西银行在巴西境内有近4万家网点，业务涉及银行业的所有领域。服务的行业主要包括：农业，飞机，通讯，[计算机](http://baike.baidu.com/view/3314.htm)，基础建设，数据处理，工业医药，石油等。巴西银行同时是一家国际化银行，其由于准确的市场把握能力使得其业务范围遍及世界40多个国家，拥有庞大的海外分支机构网，在所有的世界金融中心都有巴西银行的身影。

信用卡业务是巴西银行主要的零售业务之一，也是其个人业务的主要利润来源之一。但自2008年国际金融危机之后，信用卡业务的违约率大幅上升，从而对信用卡业务的利润带来严重威胁。2014年1月份，信用卡风险控制部门（以下简称“部门”）决定开发一张信用卡的申请评分卡，从而更好的控制信用卡业务的风险。评分卡建成之后将在该行的信用卡审批系统NBSM中嵌入实施。

**二、数据描述**

经过预分析，部门决定选取2011年10月份到12月份这3个月申请信用卡的用户用来建模。在这三个月中共有150123人申请巴西银行的信用卡，其中2345人中途放弃了申请，57283人被拒绝，剩下90495人获得批准。表现窗口为24个月，“坏”客户定义在在表现窗口中出现过90天及以上的逾期（Vintage Curve和Roll Over Rate的数据见bad definition.xls）。有394个客户在表现窗口内丢卡或者死亡，因此剩下90101个客户具有表现信息。

用来建模的数据主要分为两部分，一部分是客户申请是提供的一些基本信息（性别、年龄、收入等等），数据见modeling\_demo.sas7bdat。另一部分是巴西银行从巴西信用局获取的客户在申请时的信用记录（贷款和信用卡总数、债务比、过去两年的逾期记录等等），数据见modeling\_buro.sas7bdat。具体的变量解释见Data Dictionary.xls。此外，2012年1月份的数据被用来作为验证样本，具体数据见validation.sas7bdat。2014年1月份的数据被用来做PSI检验，具体数据见psi.sas7bdat。

**三、项目要求**

每个小组独立根据上述背景和提供的数据完成一张申请评分卡的建模，最终提交的文件包括：1、SAS程序（.sas）。2、working sheet（excel）。3、建模文档（word）。4、汇报的ppt。最终还要推选一名代表在课堂上对本小组的建模进行汇报。

在项目进行过程中，4月11日上午各个小组分别汇报项目进度情况（汇报时要求完成项目计划书，建模工作进行至粗分区）。从8：00开始，每个小组20分钟。地点：办公楼1102。

4月20日24：00之前提交最终的全部文件（SAS程序、working sheet、建模文档、汇报的ppt）。4月23日课上进行汇报。每个小组10到15分钟。

**四、建模文档所包含的内容**

第一部分：项目概述，包括：项目背景、商业目标、业务机会、实施策略等。

第二部分：数据和变量，包括：数据来源、样本范围、观测点的选择、表现窗口、验证样本和PSI样本的选择、数据流、Vintage Curve、Roll Over Rate、因变量的定义（包括建模样本和验证样本的坏账率）、具体的建模变量等。

第三部分：评分卡模型，包括：最终的评分卡、模型验证（Lift Plot, Score distribution, log-odds）以及其他一些可能需要说明的问题。

第四部分：团队成员和各自的分工。