

美国 Prosper 借贷平台贷款可视化分析

这个 Tableau 分析数据集选自 Prosper 公司 2005-2014 年间的贷款数据，它一共包含了 81 个变量，113937 行贷款数据，包含了贷款数量，借款者收入，职业，借贷利率，信用等级等指标。这个 Tableau 分析只选取了一部分感兴趣的变量进行分析。

总结

这个 Tableau 可视化主要想探索在 2005-2014 年间贷款的变化情况，和在经历了金融危机后，Prosper 贷款是一个什么样的状况。首先，我想看这些贷款数量在不同州的基本分布情况以及当前拖欠额度在不同州的分布情况。在知道了贷款的分布以后，还想进一步探索一下 2005-2014 年总体时间区间内贷款状态的分布，接下来我探索了职业群体贷款数量分布，人们贷款的目的，和最多的贷款目的中的职业数量占比，还有在不良贷款情况下，职业群体数量，人们贷款的目的，最后，从拖欠和信用等级良好贷款和不良贷款角度来分析金融危机以后贷款的表现在这几年间的变化情况，是变好了还是变差了。从可视化中得出的结论是：1.贷款数量和拖欠额度最多的州都是一样的，可能贷款数量和拖欠额度之间存在一定关联；2.从 2005-2014 年贷款状态的分布来看不良贷款越来越少；3.其他、专业人士的借款数量和贷款状态为不良贷款的借款数量都最多，人们借款的目的最多是为了债务合并，而借款目的为债务合并中最多的职业群体是其他、专业人士；4.从拖欠和信用等级的角度来看 prosper 贷款金融危机后的贷款表现都有一定的改善。

我的第一个故事版本：

https://public.tableau.com/profile/fairy#!/vizhome/prosperloanTableau_15522746443520/sheet0 （prosper 借贷平台贷款分析最初版）

反馈：

根据反馈，主要的问题有：

1.整体问题

- 1.1. 分故事太多，每个故事的内容比较少，2-5 个比较合适，将相似的合并起来。
- 1.2. 每个分故事的数据图旁边没有解释说明、分析过程，开头应该放引导语或者结论，结论也可以放到末尾。
- 1.3. 分故事的标题太长，应只描述主题而不是结论
- 1.4. 分故事之间最好是递进、总分或者并列关系，选择一种，然后体现在分故事的标题中和引导语中。
- 1.5. 本项目的目的是讲好一个故事，并用数据支持你的故事。故事自成体系即可，对数据可以有取舍，数据可视化也要围绕你的故事。

2. 细节问题

分故事 1

1. 图 2 color bar 用单色，因为数据都是正数，是一个方向的，用红蓝两色没有意义。
2. 图 1 不错。
3. 图 1 和图 2 可以合并一下。
4. 图 2 对应的筛选器可以隐藏掉，年份的那两个，比较不需要互动。

分故事 2

1. 一些项目的数量太少，可以合并起来，比如 Past due xxx 可以合并成一个，因为在这张图中，past due 多久不重要。
2. 筛选器 发起年这个，可以隐藏起来。
3. 故事内容太少，考虑扩充或者合并。
4. 表 1 的 other 很扎眼，肯定不能这么放着，要么解释 other 是什么，要么删掉。
5. 还是筛选器，不用的话就隐藏掉。

分故事 4

类似分故事 3。

分故事 5

1. 尤其体现了只有数据，没有故事，现在只是描述了图表本身，但这个图表表示了美国发生了什么呢？如何证明的？需要加一些说明的文字。其他分故事也是如此
2. 图 2 缺了 2005、2006 年的绿色数据点。
3. 筛选器应该放在对应的图表旁边，拖到图 2 旁边。

分故事 6

1. 隐藏筛选器。
2. 两个 color bar 都不合适。
3. 可以用直方图表示百分比变化，横轴是时间，纵轴是百分比，配合表格。
4. 2005-2008 的数据既然没有，直接用文字说明就可以了，不用体现到图表里。

设计与反馈后的修改

我设计故事的想法是，我要叙述的是这样一个故事：先从总体时间区间开始，从地图、贷款状态先来看 prosper 贷款的总体情况，再来看看不同职业群体的借款目的，和借款状态为违约贷款的职业群体和目的，最后再来更细地看金融危机后贷款的表现是怎样的，是变好了还是变好了，最后的这一部分我想到了用拖欠是否严重和信用等级与违约贷款来看贷款的表现状况。

根据反馈，做了如下修改：

整体方面：每个故事的标题部分都围绕主题取了一个标题，每个故事都添加了文字说明和结论。

故事细节方面：

1. 故事一中的两个地图合并为一个图了，没有用颜色突出数量的不同，而是用圆形大小体现数量的不同。我隐藏了年份筛选器。这样故事一看起来简单清楚，引入主题。
2. 故事二：我把贷款状态的所有英文类别改为了中文，同时把 Past due 很多天的分类全部归类为已违约，隐藏了年份筛选器，在直观图对应的类别部分都增加了相关的文字。这样看起来更清楚直观。我把贷款状态分成了不良贷款和良好贷款两类，不良贷款包括冲销坏账、已到期、已违约，良好贷款包括当前贷款，已完成，最终付款。我加了良好贷款和不良贷款两个折线图可以更好地看出良好贷款和不良贷款在总体时间区间内的分布。
3. 故事三、四：我隐藏了年份筛选器，在借款目的中加入了用颜色筛选职业，这样就能看出借款目的中不同职业群体占比。
4. 故事五：我把两幅图的位置对掉了，同时把当前拖欠额度变化折线图当前的当前拖欠额度平均值换成了当前拖欠额度总和，因为平均值经过除法之后，它表示每一年的拖欠数量就不是很准确，而用总和来表示就相对更准确。我去掉了 2005 年-2008 年，因为我主要想知道金融危机以后的贷款表现，所以这几年我不用看。

5. 故事六：我把两个表格都换成了直方图列百分比，同时我把原先的违约
贷款信用等级数量和信用等级数量图换成了不良贷款和良好贷款信用等级
数量分布图，去掉了 2005 年-2008 年。

我的最终故事版本：

https://public.tableau.com/profile/fairy#!/vizhome/prosper_1527262401390/sheet0?publish=yes (prosper 借贷平台贷款分析最终版)

参考资源：

https://blog.csdn.net/Sundy_L/article/details/80890395

https://onlinehelp.tableau.com/current/pro/desktop/zh-cn/calculations_calculatedfields_formulas.htm

https://onlinehelp.tableau.com/current/pro/desktop/zh-cn/actions_highlights_colorlegend.htm