**Forecast Rossmann Store Sales**

(from Kaggle Competition)

**题目描述**

[](https://github.com/nd009/capstone/blob/master/Rossmann_Store_Sales/rossmann_banner2.png)

Rossmann是欧洲的一家连锁药店。 在这个源自Kaggle比赛[Rossmann Store Sales](https://www.kaggle.com/c/rossmann-store-sales)中，我们需要根据Rossmann药妆店的信息（比如促销，竞争对手，节假日）以及在过去的销售情况，来预测Rossmann未来的销售额。

**数据下载**

此数据集可以从Kaggle上[下载](https://www.kaggle.com/c/rossmann-store-sales/data)。

**建议**

对于数据的使用，可以有很多做法，比如说，可以模拟参加比赛，用train.csv来建模，之后用test.csv来预测，通过提交到Kaggle来评估模型表现。

另一种做法，可以只是使用train.csv, 进行本地的分割，预留一部分作为test data, 其余用来建模。这样就不需要提交去Kaggle， 也可以自己重新定义指标来衡量模型的表现。

另外，也可以只是借助数据，来自己设计问题（比如原题目中对销售额的预测是一个regression问题，这里也可以转化成对Rossmann药妆店进行分类）。

无论使用什么方法及获得什么样的结果，值得注意的是，这个项目中会遇到很多数据科学常见的问题，比如缺失数据该如何处理，有时间序列的数据该如何分割，如何利用时间信息，如何使用store.csv提供的额外信息，等等。这些问题可能需要更进一步的思考，这些思考的过程，最好写进最后的报告中。

当然，如果想要简化问题，可以忽略一些相关的信息/特征。项目的基本要求，可以参见Udacity的[项目要求](https://review.udacity.com/#!/rubrics/273/view)。

**提交**

* PDF 报告文件（注意这不应该是notebook的导出，请按照[模板](https://github.com/nd009/machine-learning/blob/zh-cn/projects/capstone/capstone_report_template.md)填写）
* 项目相关代码
* 包含使用的库，机器硬件，机器操作系统，训练时间等数据的 README 文档

**参考**

比赛第一名的[采访](http://blog.kaggle.com/2015/12/21/rossmann-store-sales-winners-interview-1st-place-gert/)及[参考资料](https://www.kaggle.com/c/rossmann-store-sales/forums/t/18024/model-documentation-1st-place)。