

钻石价格预测

阮祥炬

2019/6

Step1:项目概述

一家钻石经销商最近决定退出市场，库存有一套 3000 颗钻石供拍卖。看到这是扩大库存的绝佳机会，一家珠宝公司已表示有兴趣参与竞标。为了确定出价多少，该公司的分析团队使用庞大的钻石价格数据库（50000 条）来建立线性回归模型，以根据其属性预测钻石的价格。作为业务分析师，您的任务是应用该模型来推荐公司应该为整套 3,000 颗钻石投标多少。

Step2:模型理解

钻石价格的模型可以表示为克拉数（carat），切割等级（cut），净度（clarity）这几个影响因素的方程。

经过对 50000 条原始数据进行线性回归分析得到各影响因素的值，得到如下公式

价格 = -5,255 + 8,363 x **克拉** + 160 x **切割** + 458 x **净度**

1. 根据这个模型，如果有一颗钻石比另一颗钻石重 1 克拉，切割等级和净度一致，我应该多付多少钱？为什么？

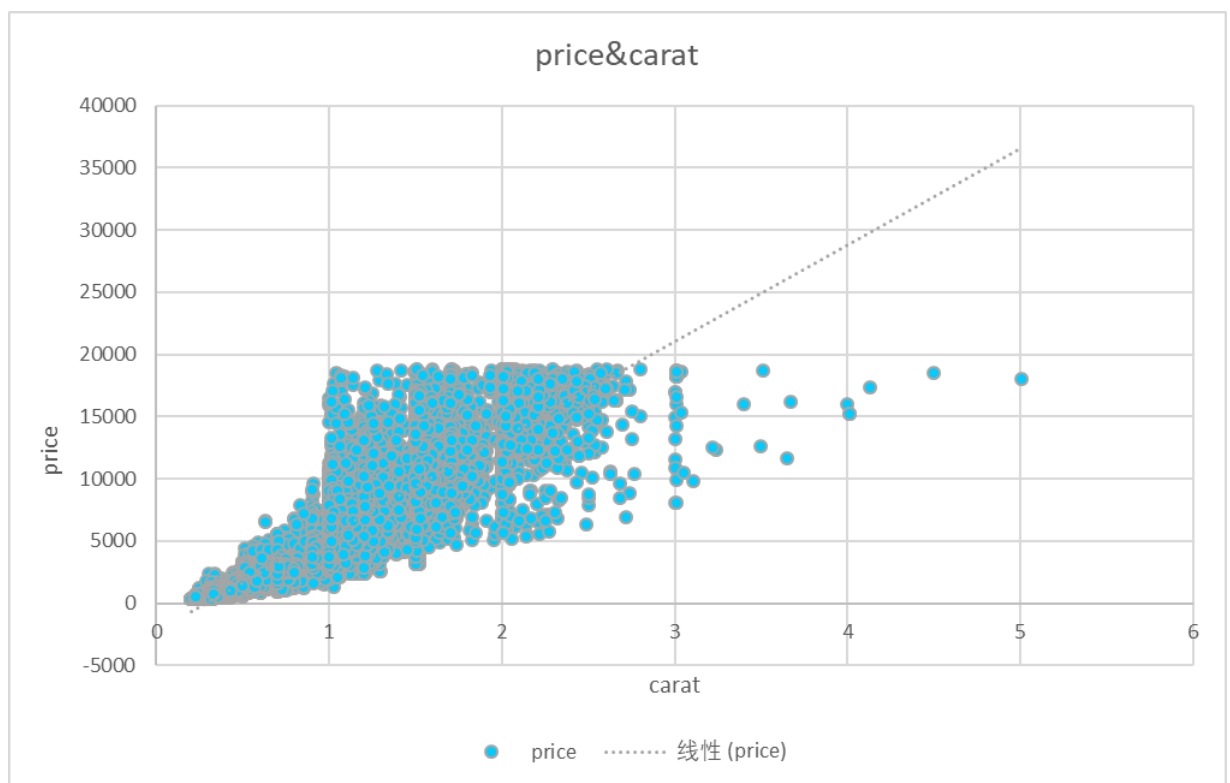
根据这个模型，在切割等级和净度一致的情况下，一颗比另一颗重一克拉我们希望付\$8363。

2. 根据这个模型如果你对一颗 1.5 克拉，切割等级为 very good（在此模型中的等级使用数字表示为 3），净度为 VS2（在此模型中的等级使用数字表示为 5）的钻石感兴趣，这个模型预期的价格是多少？

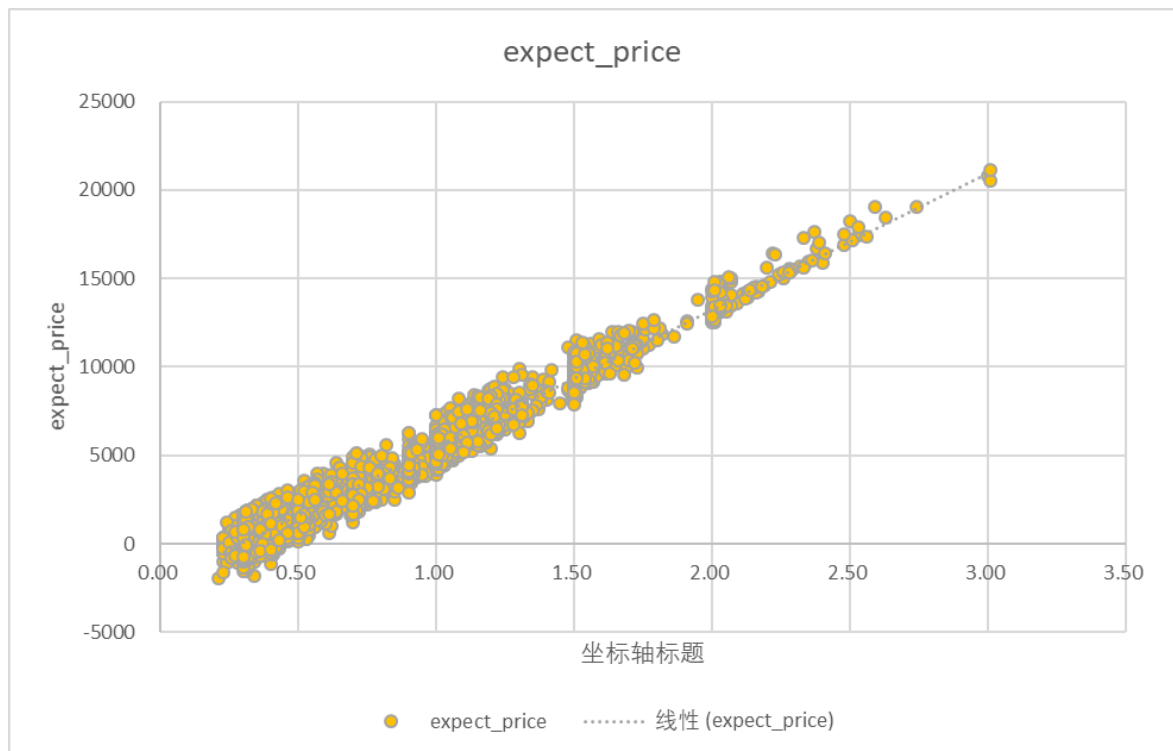
这个规格的钻石，根据给出的价格预测模型判断，预期的价格为\$10059。

Step3:数据可视化

1. 图中展示的是数据库中的钻石的数据，x 轴表示钻石的克拉数，y 轴表示钻石的价格。



2. 图中的数据表示的是预测的钻石的数据，x 轴表示钻石的克拉数，y 轴表示钻石的预测价格。



Step4:推荐

1. 你推荐公司决定以什么价格购买这批钻石？请解释如何得到这个数字的。

通过使用原始预测，可以算出这批钻石的总价格为\$11727918。但是观察数据可以看出，钻石的价格有负值，这与事实不符，所以将这部分的钻石的价格调整为\$0，这时可以得到钻石的总价格为\$11859726。如果珠宝公司的目标是以零售价格的 70%向经销商购买钻石，则建议购买钻石的价格为\$8301808。