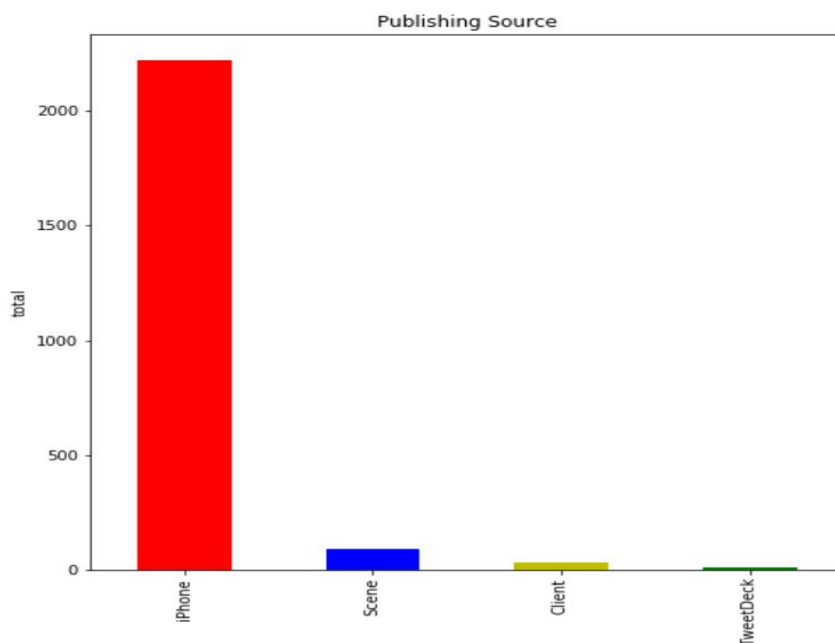


## wrangle 项目数据整理外部报告

整理（以及分析和可视化）的数据集是推特用户 `@dog_rates` 的档案，推特昵称为 `WeRateDogs`。`WeRateDogs` 是一个推特主，他以诙谐幽默的方式对人们的宠物狗评分。这些评分通常以 10 作为分母。但是分子则一般大于 10：11/10、12/10、13/10 等等。为什么会有这样的评分？因为 "They're good dogs Brent." `WeRateDogs` 拥有四百多万关注者，曾受到国际媒体的报道。

我们知道，现在是智能手机的天下，`WeRateDog` 发推特可能有多钟不同的平台，我想看看 `WeRateDogs` 最常用的发布推特的平台，我对发布推特的 `source` 数据统计后，用发布的平台数据绘制的图表如下：



从图表中可以很显著的看出，`WeRateDog` 使用 `Iphone` 发布了大量的推特消息。

在推特信息中并没有对狗的种类说明，我们使用了神经网络通过推特中的图片识别狗的种类。令人兴奋的是神经网络识别出了大部分

狗的种类，通过对数据进行统计，得出结果 **golden retriever** 这种类型的狗可识别概率最高，**web site** 可识别概率最低，看样子分类名字也有问题。

每条推特消息都有转发和喜爱的数量，不同品种的狗应该会有不同的转发和喜爱总量，通过对数据进行分析，得出结果 **golden retriever** 这种狗是转发量最大和最受欢迎的品种。