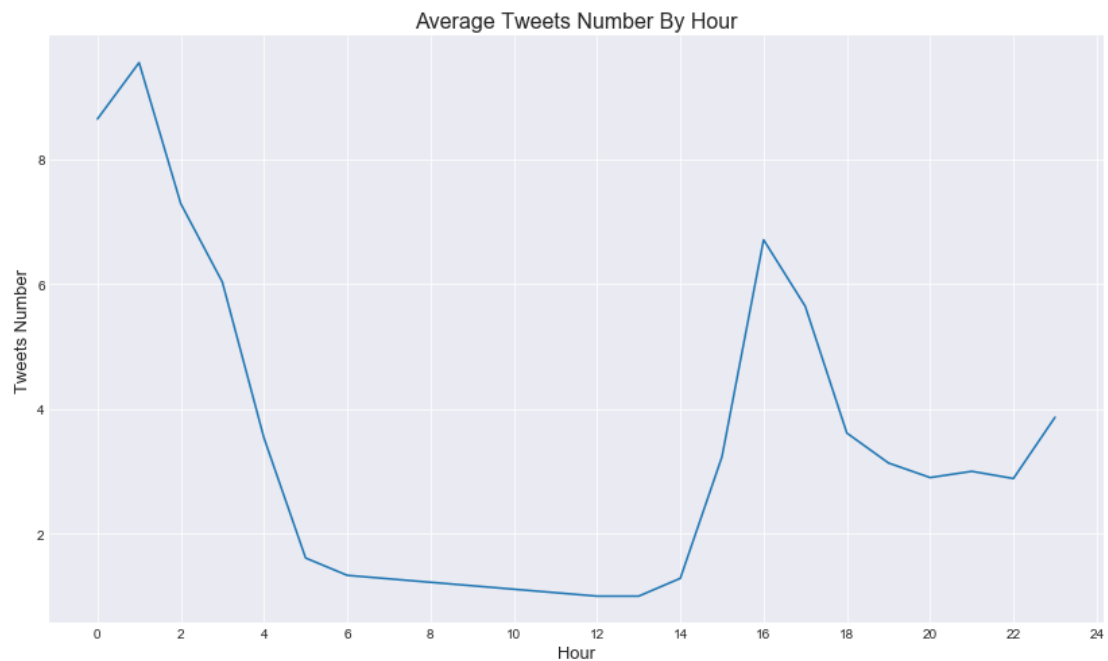


项目 8：清理与分析数据

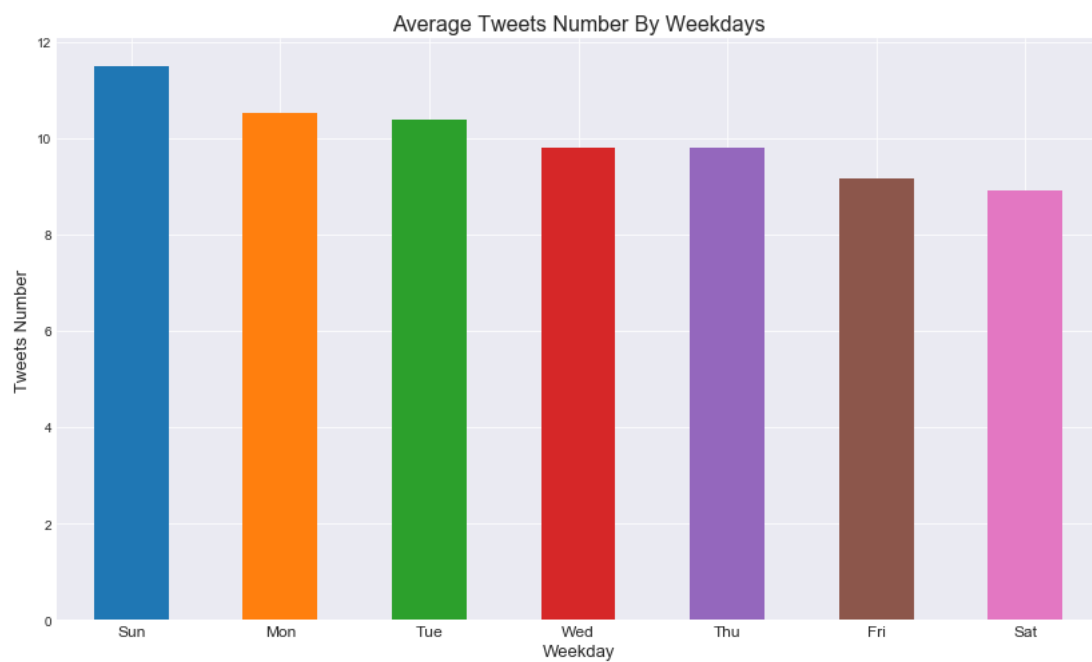
Weratedogs 推特数据分析

- 推特主的活跃时间
- 推特主的给分习惯
- 狗狗最常用的名字有哪些？
- 哪种狗狗的评分和或获赞数较高？
- 评分与获赞数、转发数有无关联？

推特主的活跃时间：



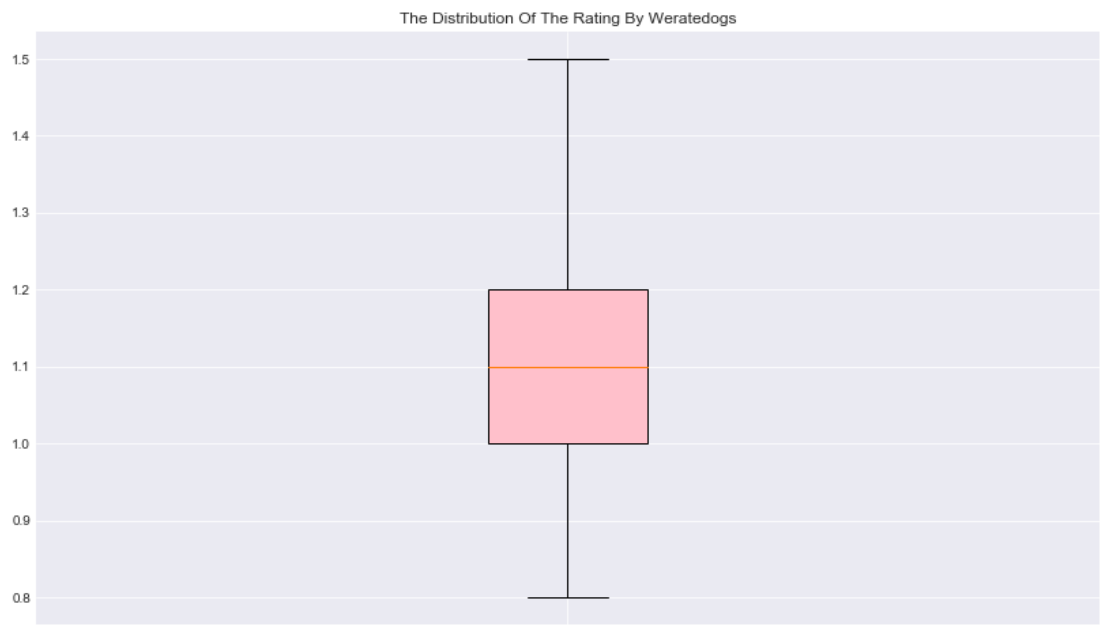
该推特主的主要活跃时间在格林威治时间(由原数据的+0000)的晚上 22 点至次日凌晨 4 点以及下午 14 点至 18 点之间，这大致对应于美国时间的上午 9-13 点及下午 17 点-晚上 23 点(以 UTC-5)，可见该推特主主要在上午和下午 17 点以后的时间比较活跃。



从星期来看，该推特主的主要活跃时间是在星期天，而从周一到周六大致是呈现下降态势。由于该推特主要是给狗狗评分，一般是回应其他推特用户的评分请求为主，因此造成这个现象的可能原因是其他用户主要是在周末发推让其评分。但这个推测是否正确需要更多数据来

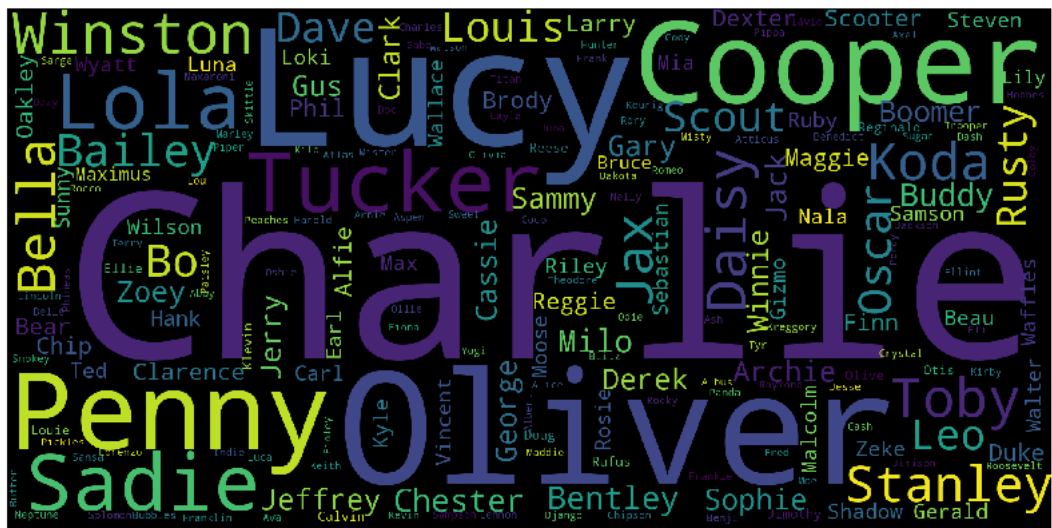
验证。

推特主的评分习惯



从上图来看，该博主最常给出的分数大约是在 1.0-1.2 之间，也就是 10/10-12/10 之间。大致上说，高于 12/10 就算是比较高的分数了，而低于 10/10 则是比较低。上图为了让图像更清晰地显示，已经去除了离群值。一般来说，打分高于 15/10 的是非常罕见的。

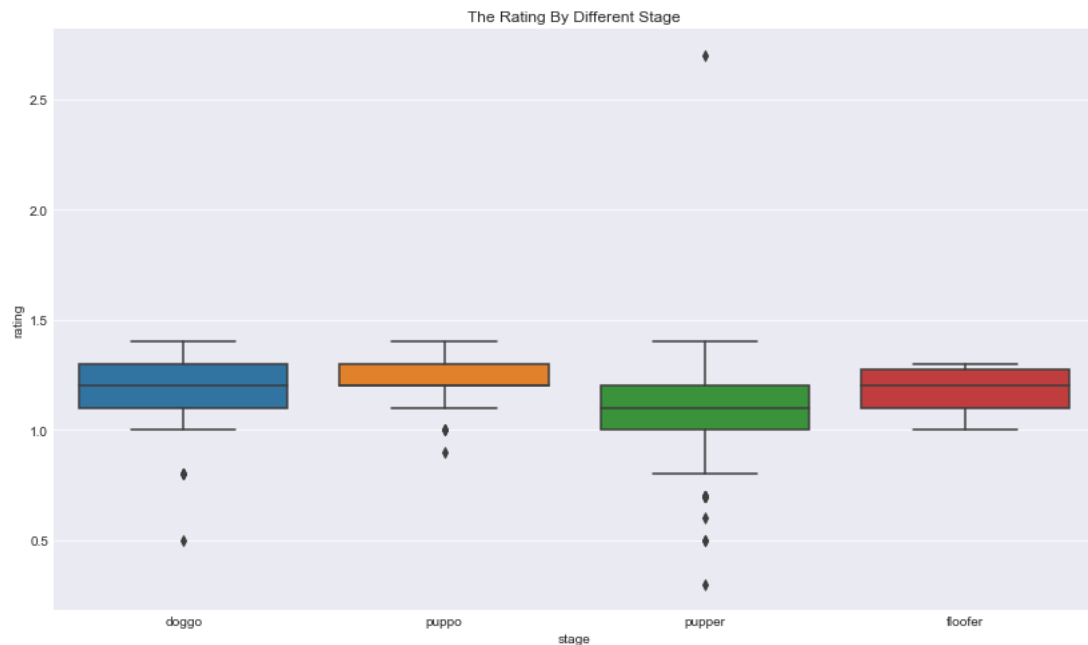
狗狗最常用的名字有哪些？



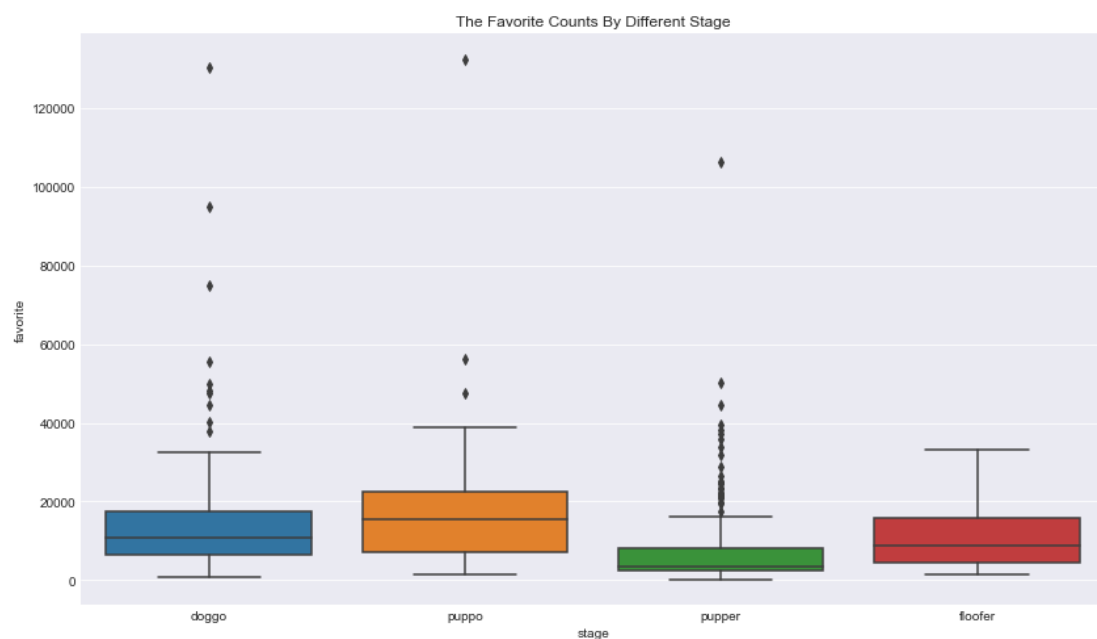
由由上图来看，狗狗最常用的名字是 Charlie、Oliver、Lucy、Lola、Copper、Lucy、Sadie

等都是狗狗比较爱用的名字。但也许这只能反应博主对狗狗的取名习惯，而不是人们喜欢给狗狗取哪种名字，因为新闻称该推特中狗狗的名字都不是真名。

哪种狗狗的评分以及获赞数最多？

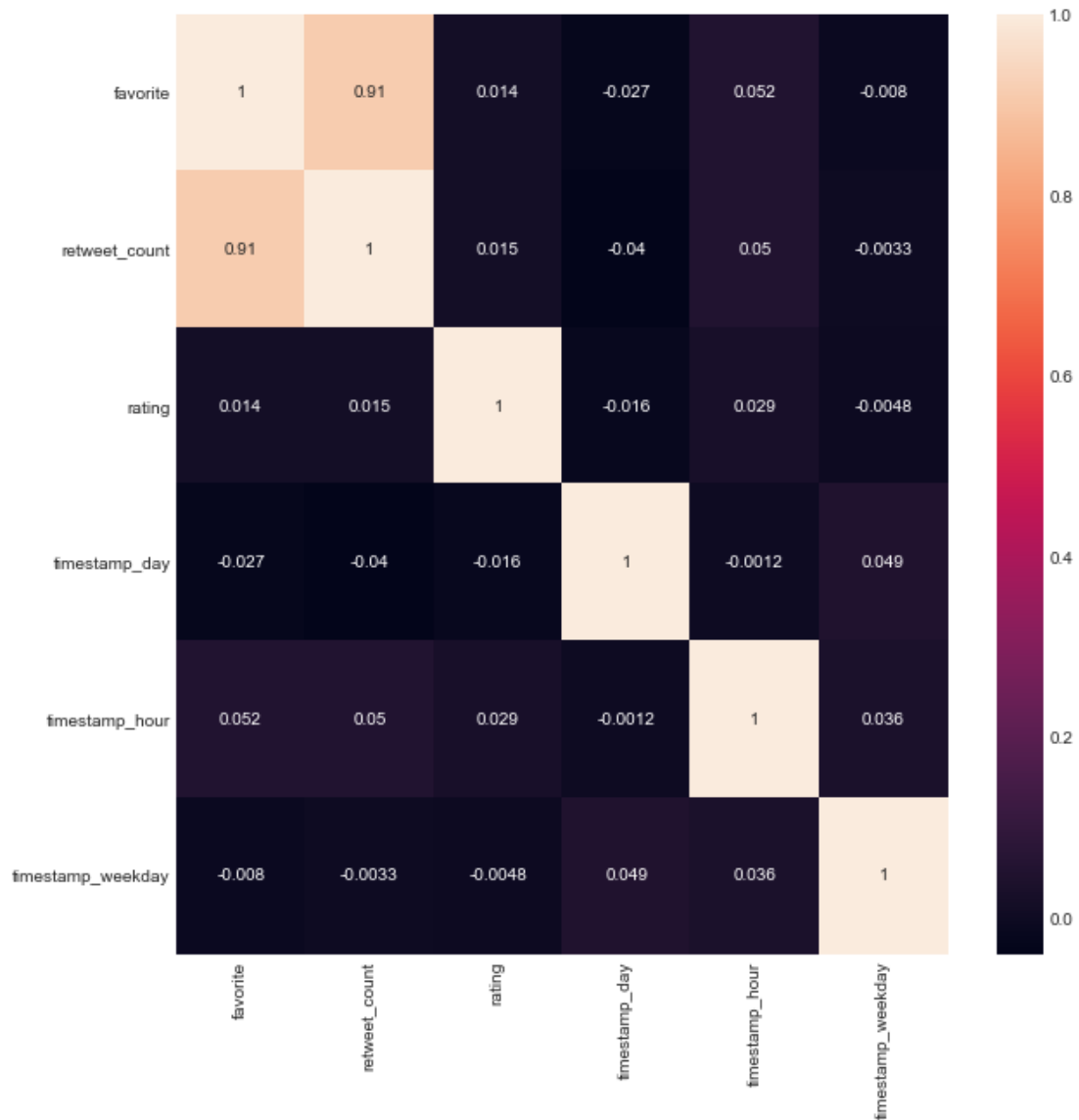


这个博主给 puppo 的分数似乎总体比其他类的狗狗更高一些，这是否说明这个博主更偏爱 puppo、更容易给 puppo 比较高的分数？由于推文中大部分并未声明其狗狗属于哪一类，因此此处能获取的数据仅仅是少数，需要更进一步使用推断统计工具后才能做出更有把握的说明。



此处同样显示 puppo 获得的赞大体上要比其他狗狗略高一些，但由于样本数量有限，同样很难就此断定人们更爱 puppo 多一点。

评分与获赞数、转发数有无关联？



我们发现评分与转发数之间并无明显的关联性，从时间上来说，评分是在转发和获赞之前，这个结果说明给出的评分更高的推特，并不见得就会相应地有更多的转发或者更多的赞（当然也不见得更少）。