> 通过对数据的评估和清理，最终得到twitter_archive_master.csv的数据，对该数据进行探索性分析

一、极值数据

通过对点赞，转发和评分进行分析，获得数据最大的狗，查看各个狗的照片。

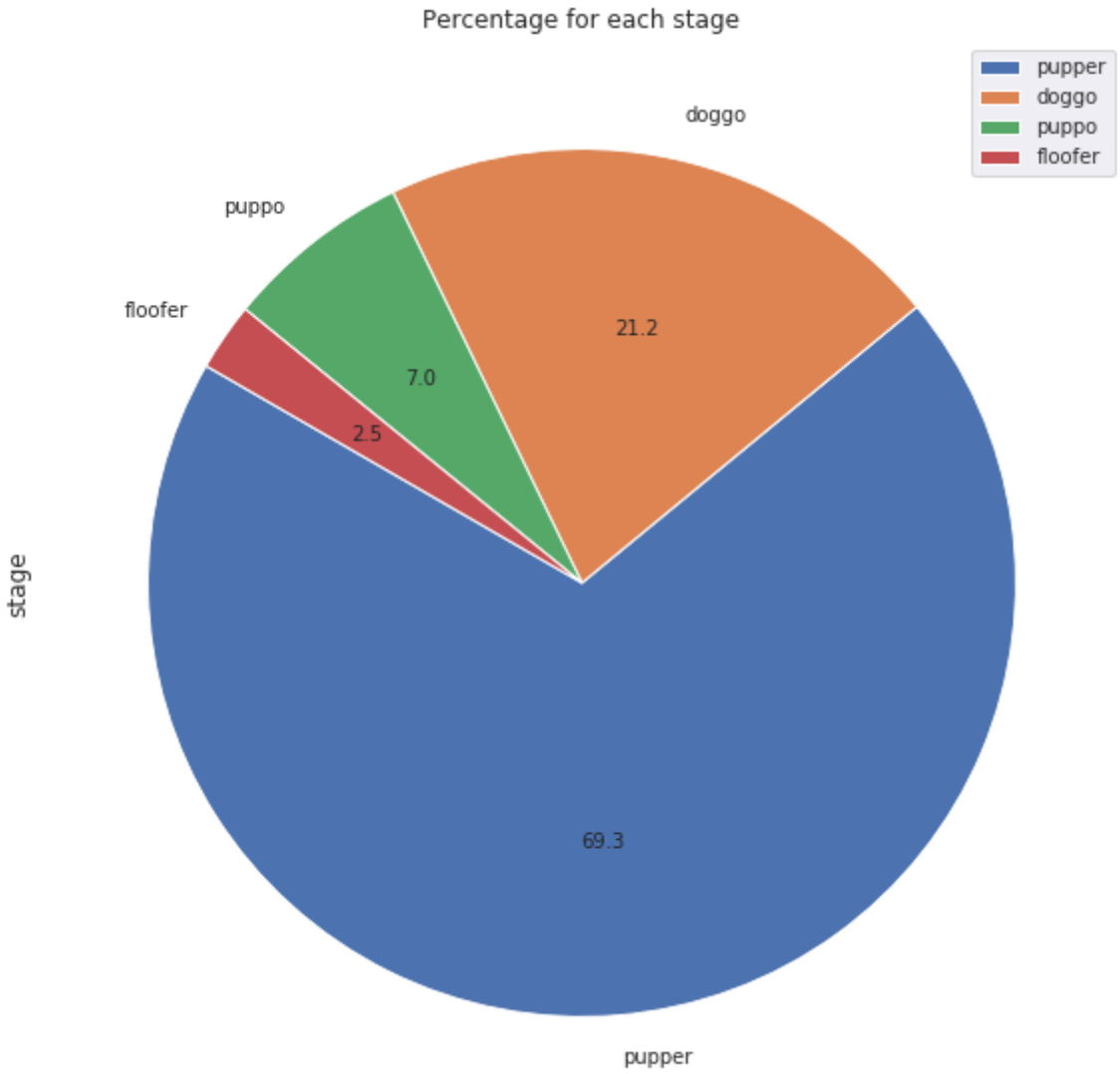
点赞最高的狗

转发最高的狗

转发评分的狗

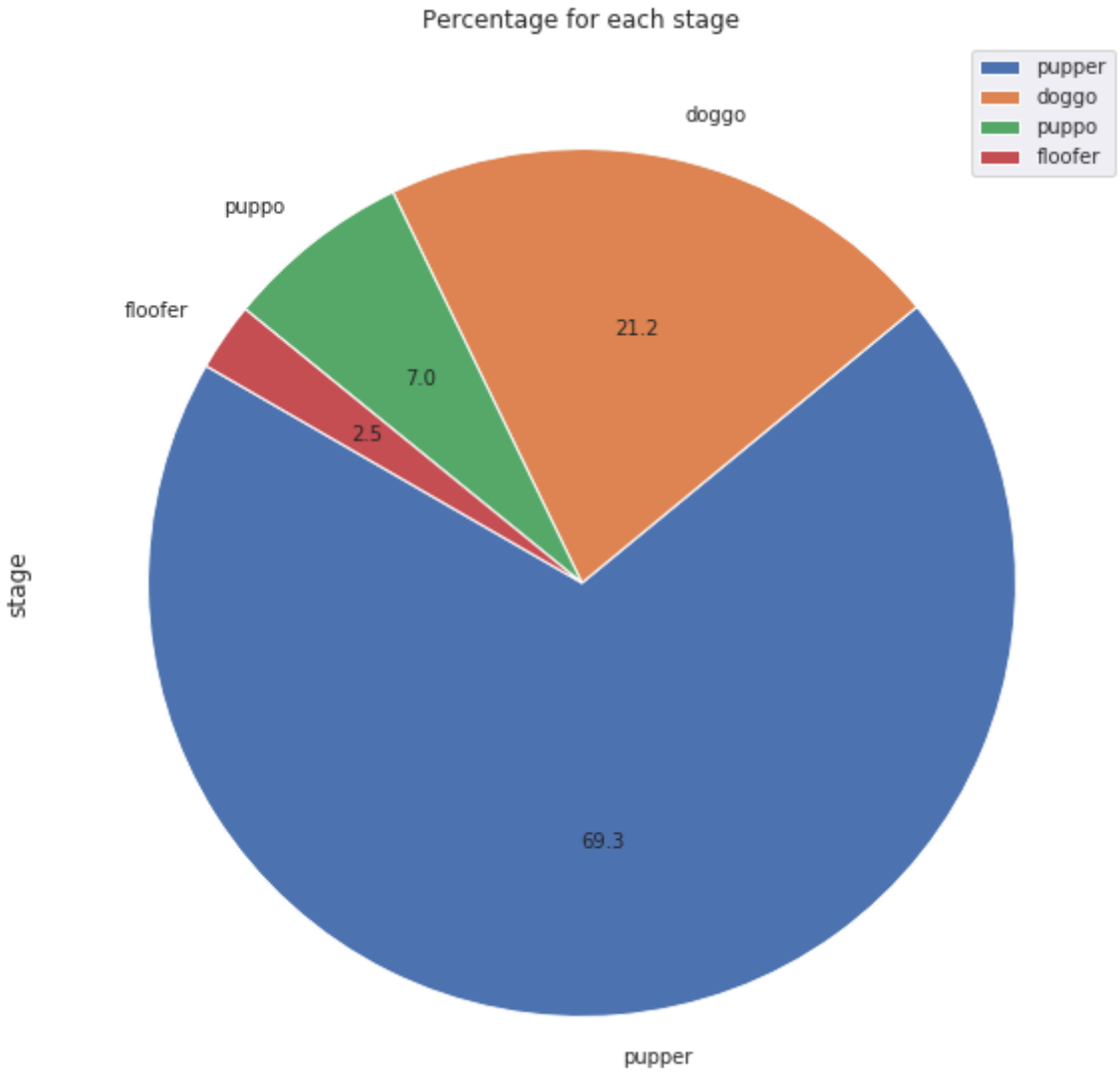
二、数据分布情况

stage分布情况



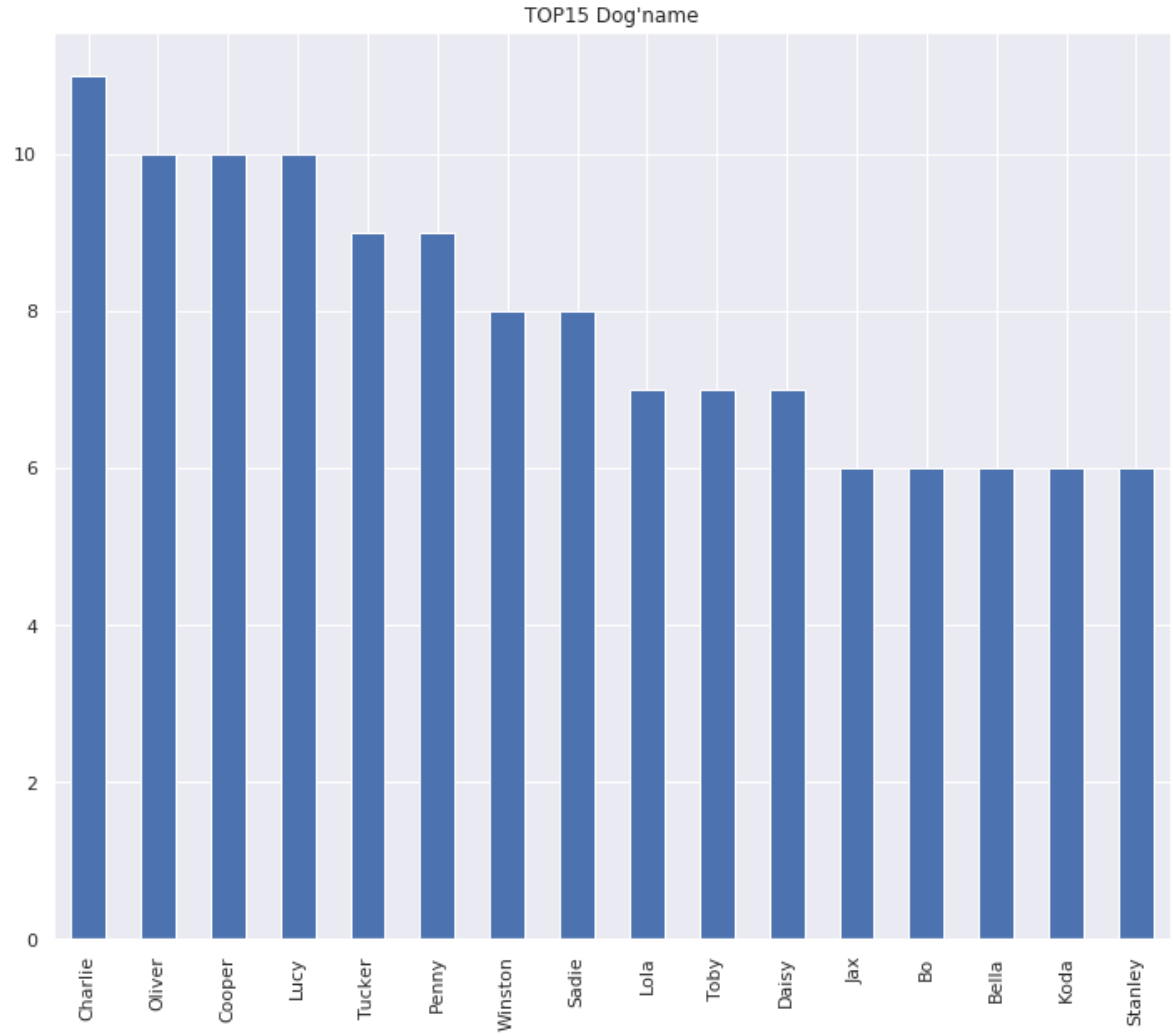
stage为pupper的狗狗占比最高

Source分布情况



- 大部分的用户喜欢使用苹果手机登录tweet
- 使用电脑客户端登录tweet的用户占比很低

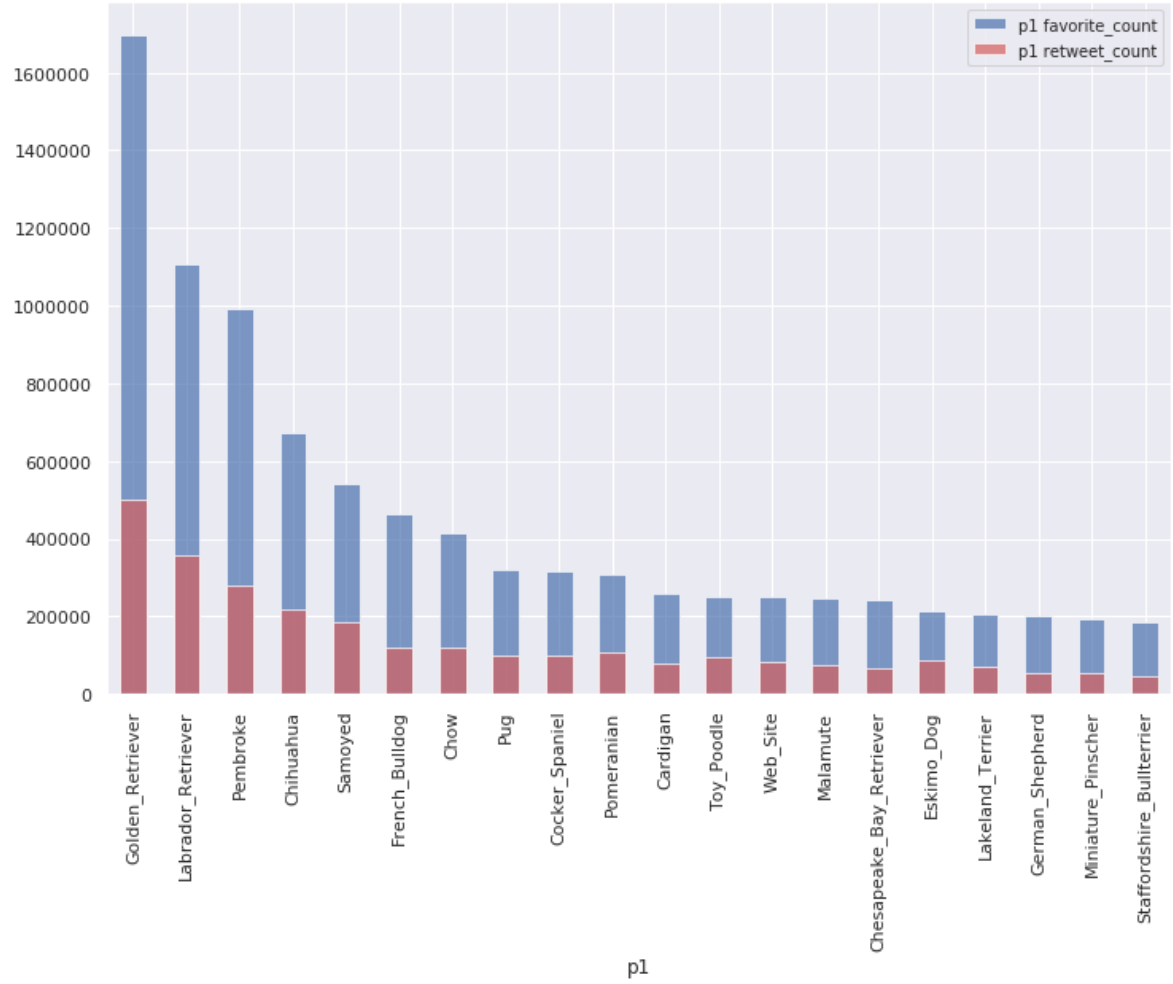
top15的狗的name分布情况



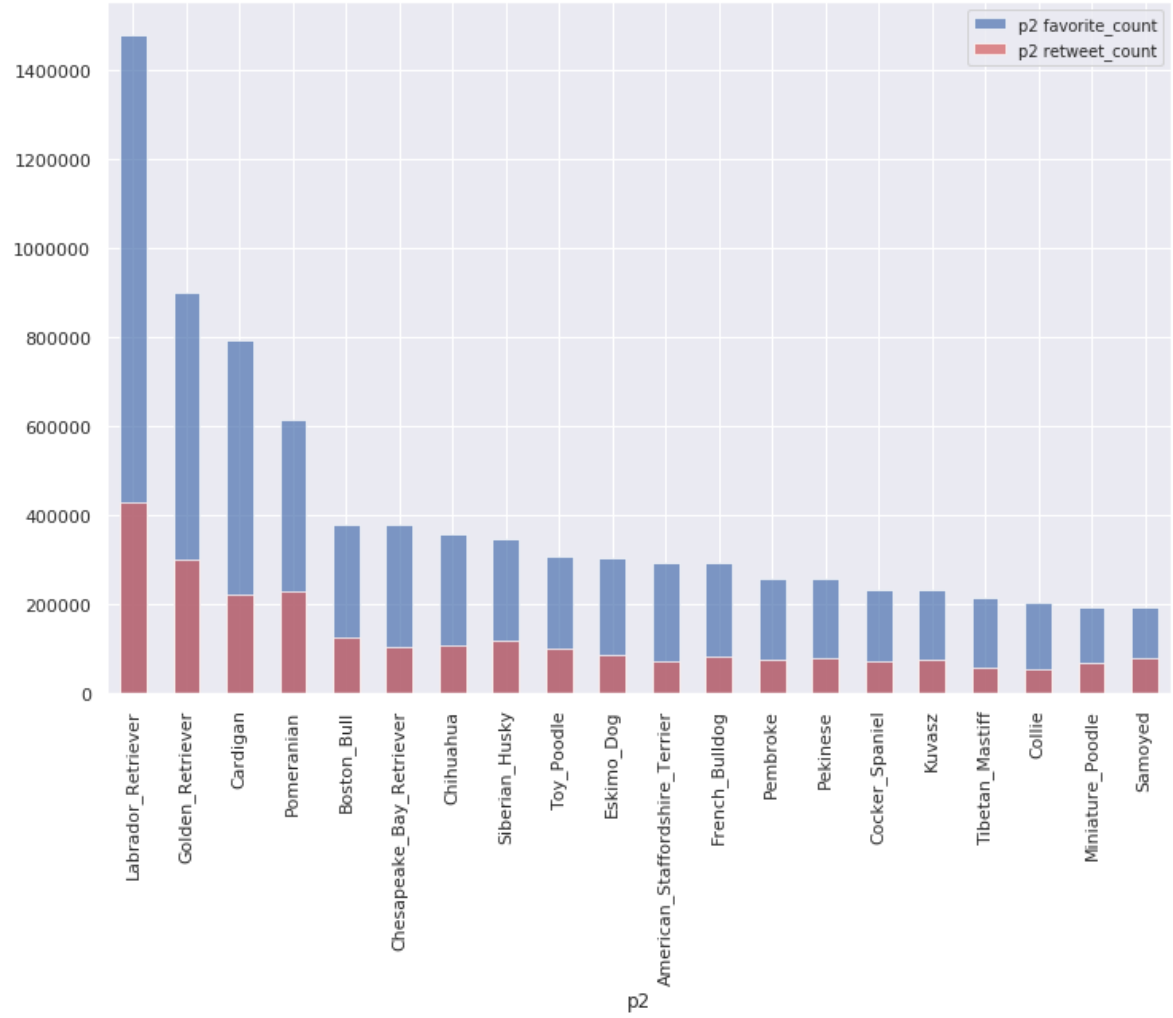
- Charlie: 这个名字使用的频率较高

狗的种类分布情况

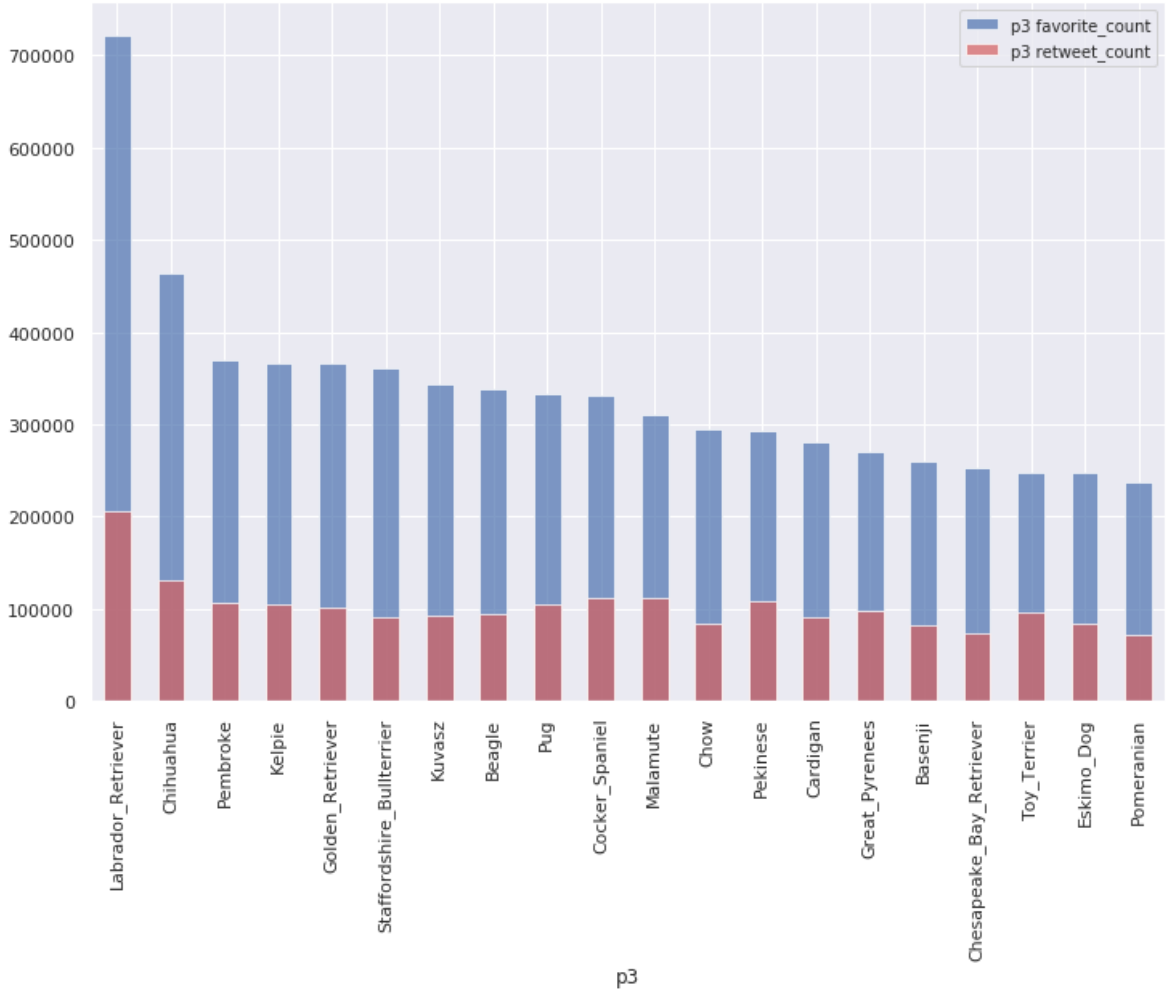
p1


p1 favorite_count
p1 retweet_count
1600000
1400000
1200000
1000000
800000
600000
400000
200000
0
Golden_Retriever
Labrador_Retriever
Pembroke
Chihuahua
Samoyed
French_Bulldog
Chow
Pug
Cocker_Spaniel
Pomeranian
Cardigan
Toy_Poodle
Web_Site
Malamute
Chesapeake_Bay_Retriever
Eskimo_Dog
Lakeland_Terrier
German_Shepherd
Miniature_Pinscher
Staffordshire_Bullterrier
p1

p2

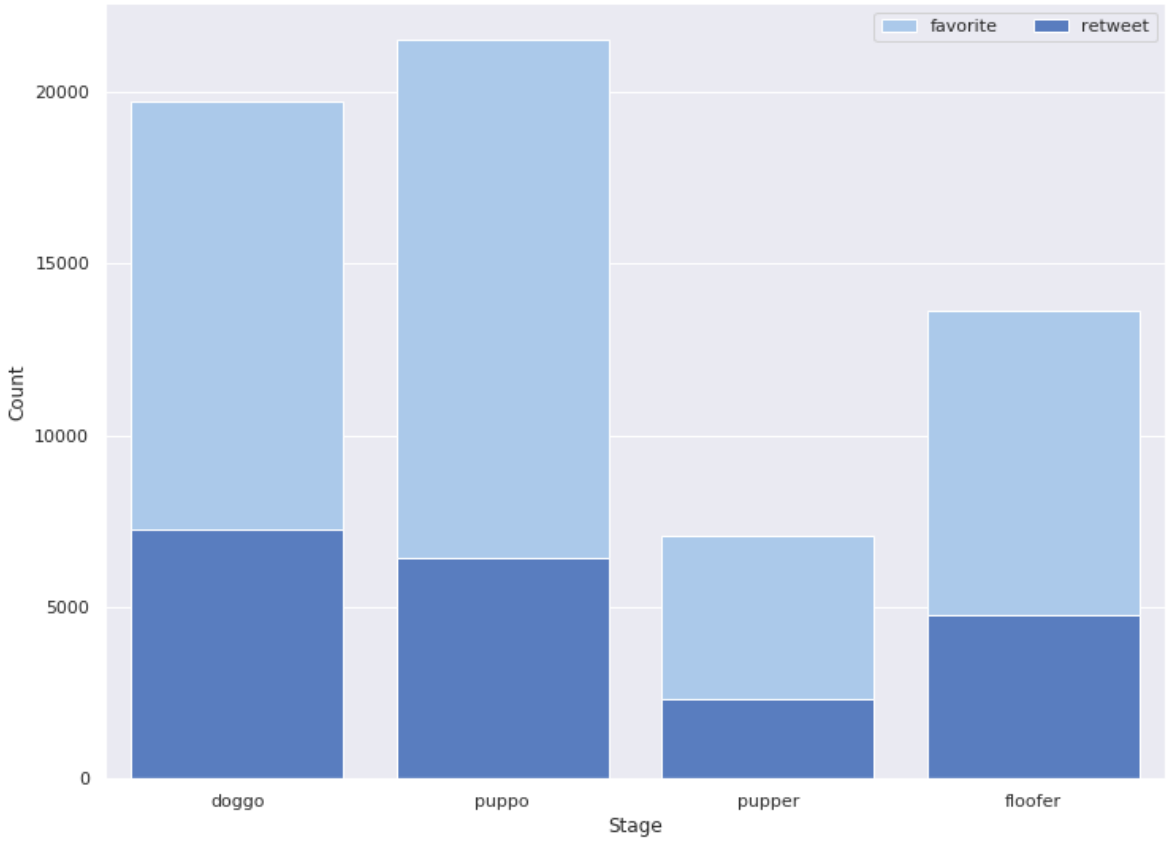

p2 favorite_count
p2 retweet_count
1400000
1200000
1000000
800000
600000
400000
200000
0
Labrador_Retriever
Golden_Retriever
Cardigan
Pomeranian
Boston_Bull
Chesapeake_Bay_Retriever
Chihuahua
Siberian_Husky
Toy_Poodle
Eskimo_Dog
American_Staffordshire_Terrier
French_Bulldog
Pembroke
Pekinese
Cocker_Spaniel
Kuvasz
Tibetan_Mastiff
Collie
Miniature_Poodle
Samoyed
p2

p3



- Golden_Retriever 在p1的识别分类中获赞和转发的数量最大
- Labrador_Retriever 在p2的识别分类中获赞和转发的数量最大
- Labrador_Retriever 在p2的识别分类中获赞和转发的数量最大

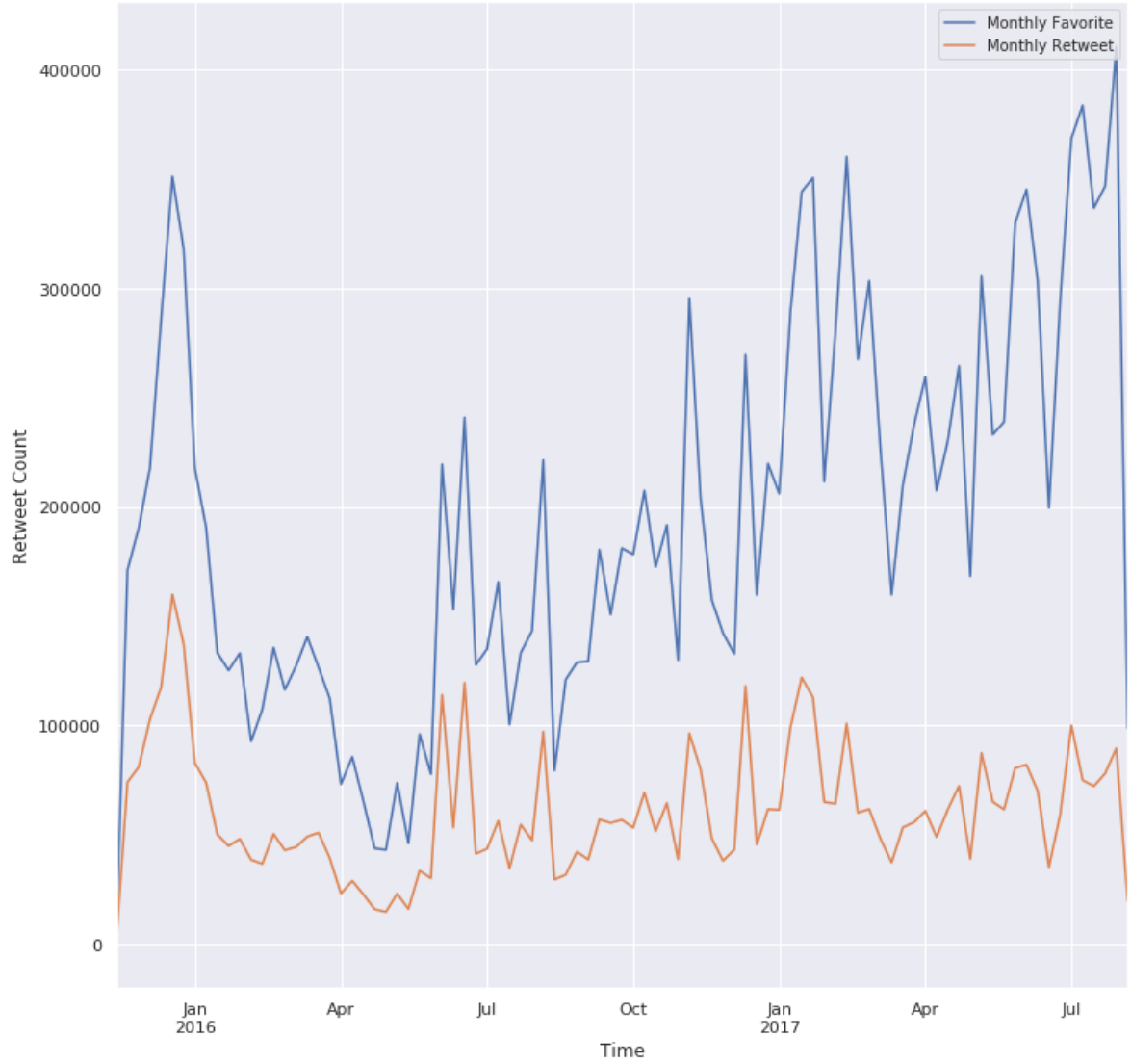
不同stage的获赞和转发分布情况



- poppo stage的狗获赞的数量最大
- pupper stage的狗获赞的数量最少
- doggo stage的狗转发的数量最多
- pupper stage的狗转发的数量最少

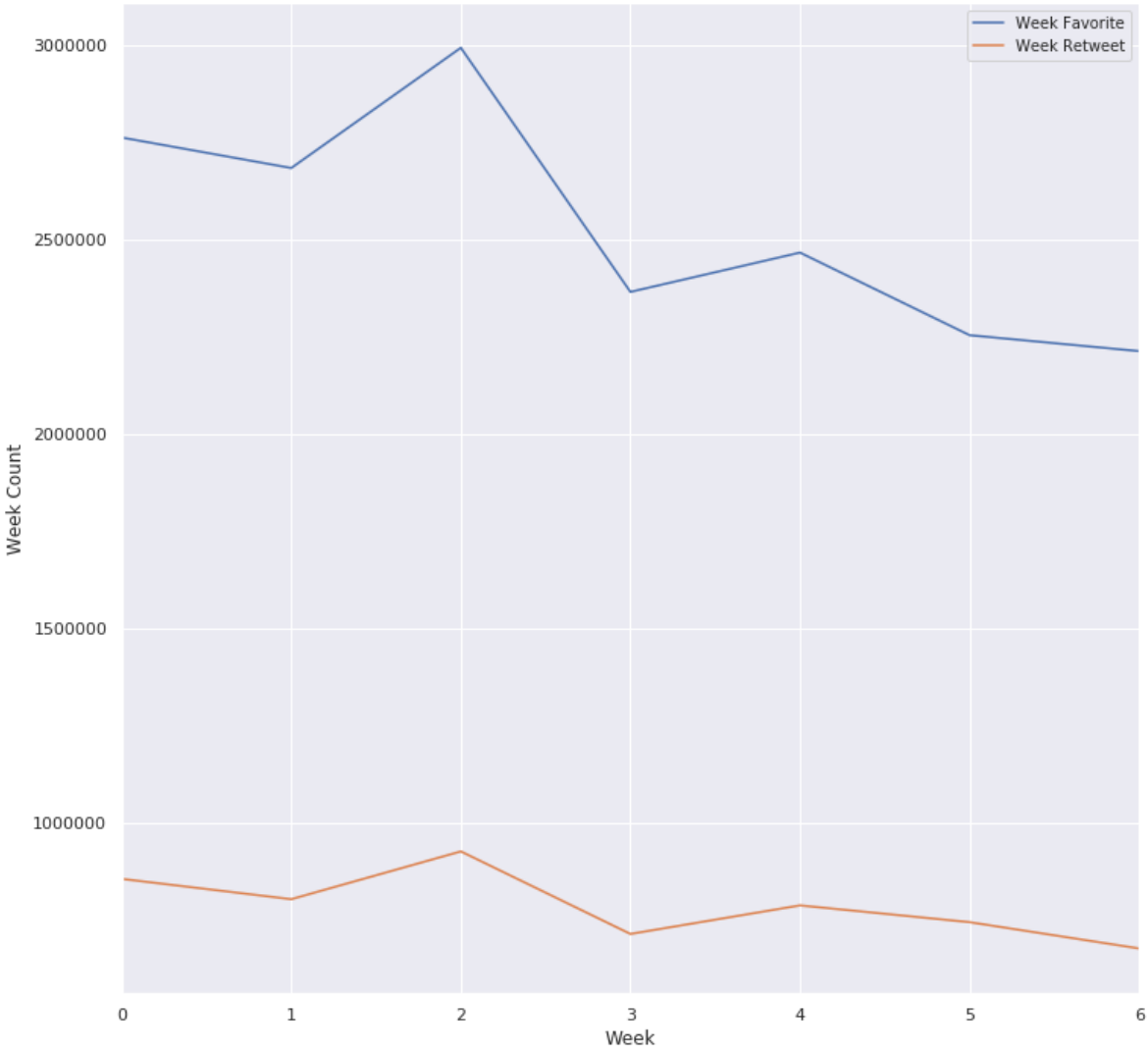
三、时间维度分布情况

月度数据分布



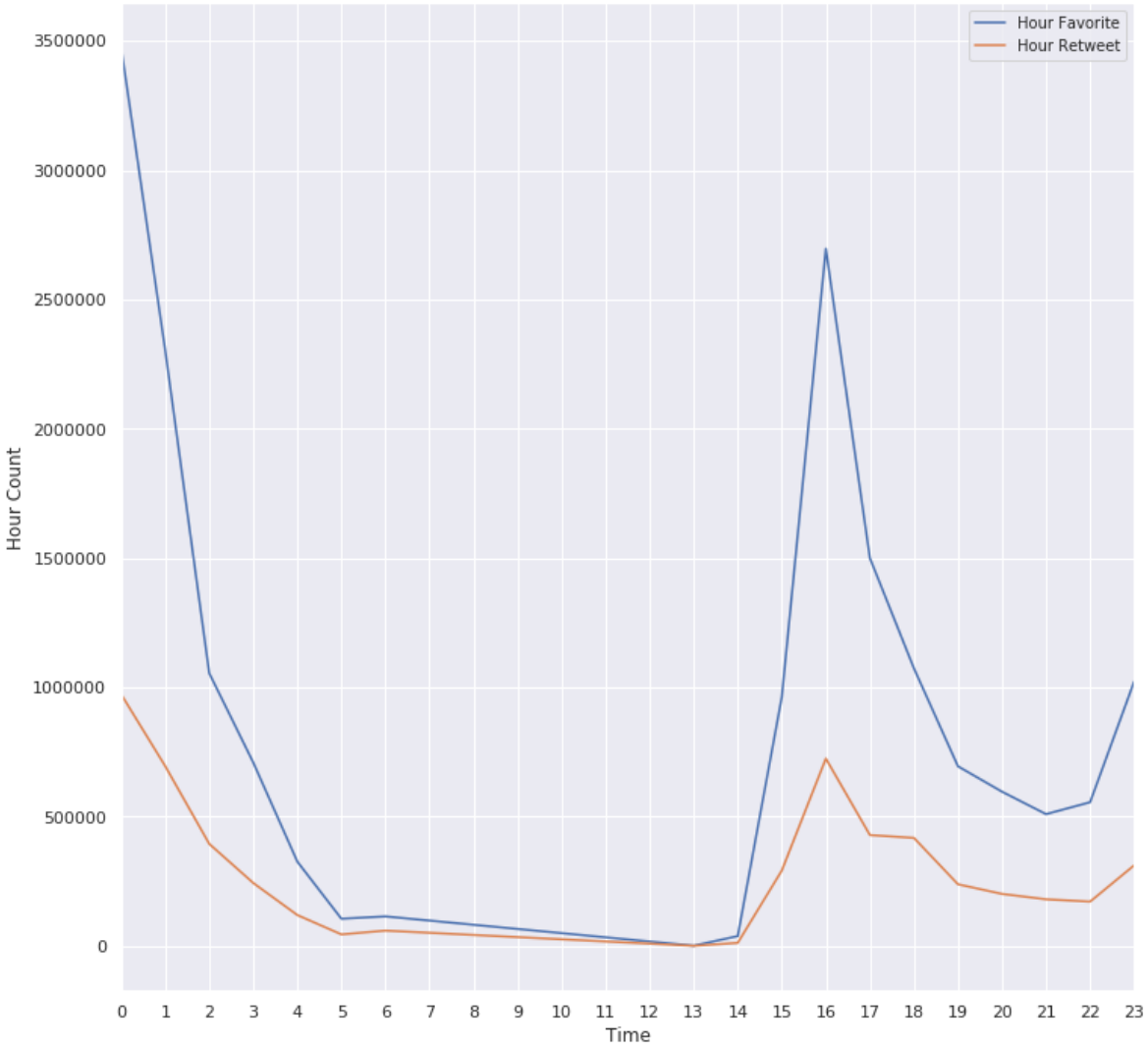
2016年年度tweet的点赞和转发开始下滑，在2016年4月份tweet的点赞和转发的数据最低

周数据分布



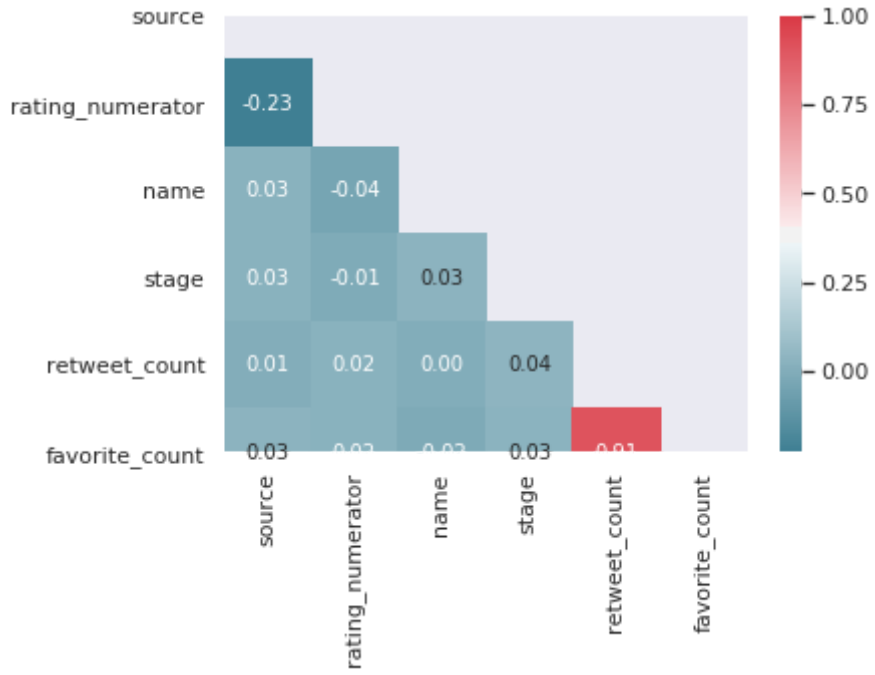
- 周二的点赞和转发数据量最高
- 周六的点赞和转发量最低

天数据分布



- 在24小时时间段内凌晨12点很多人进行tweet的操作，意味着很多人睡前都会先玩一会tweet
- 从5点到14点这个时间段，用户操作tweet的比例较少
- 在下午15到18点之间，24小时内第二个小高峰时段

四、相关性分析



etweet_count和favorite_count两者之间正向关联