

# Analyzing Reddit Engagement

...

Steven Masschelin

# Objectives

1. Identify subreddits that have high median comment and score values.
2. Determine if the subreddit itself is a good indicator of whether or not a post will land in the higher percentiles of score and comment count.

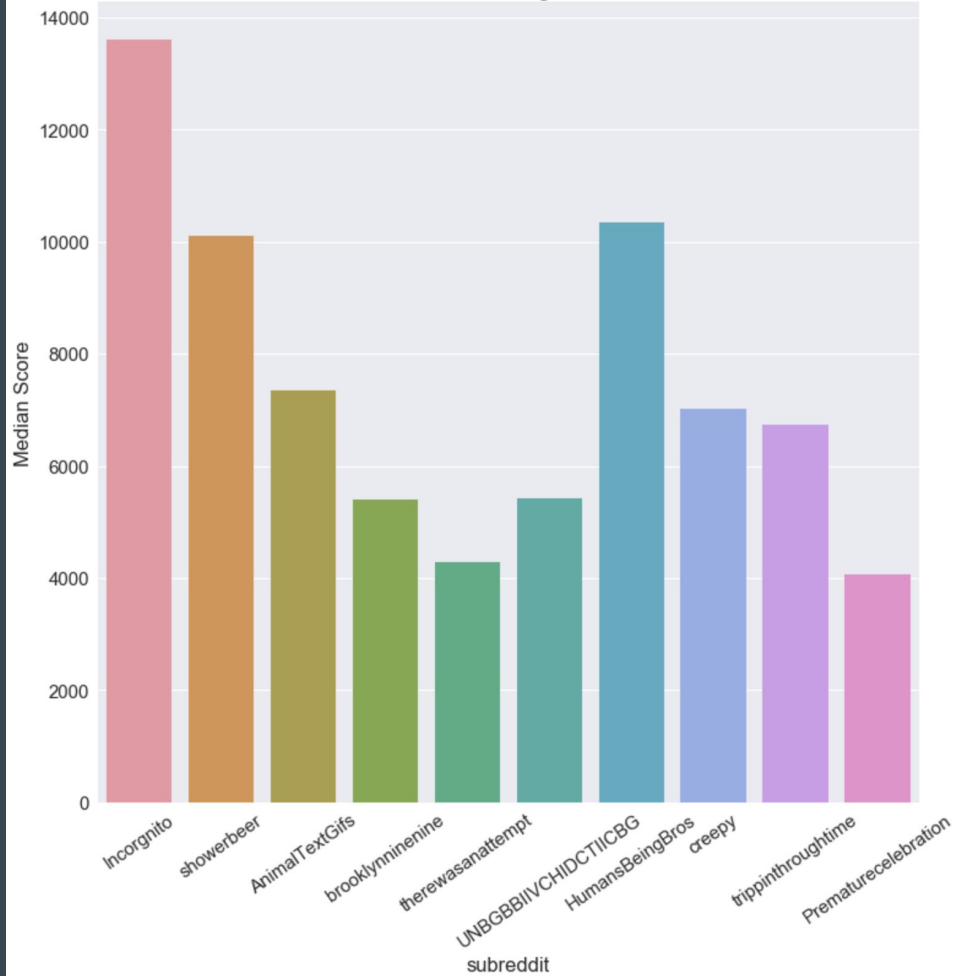
# What is Reddit?

# Methodology

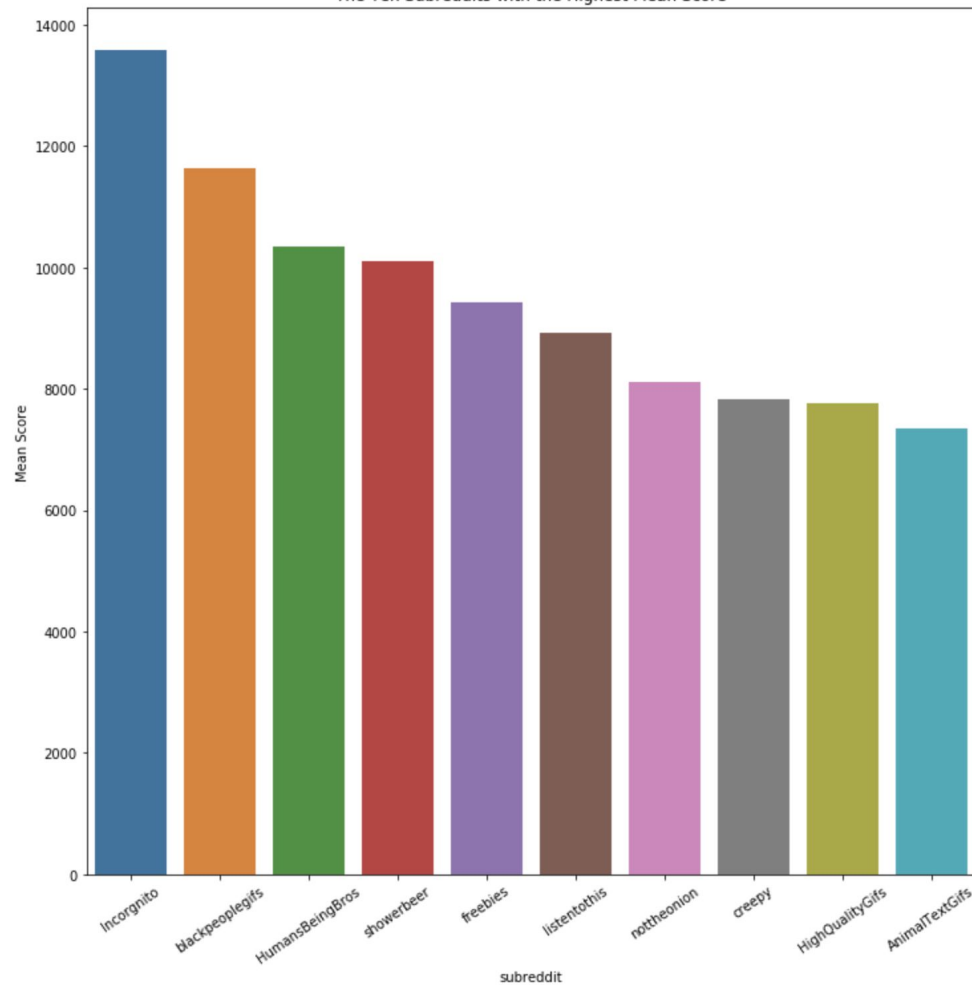
- Constructed a scraper using BeautifulSoup Package
- Data collection began at 2000 on Monday, Oct. 30 and Finished at 0000, Friday, Nov. 3rd (2017).
- Data analysis and EDA performed via Pandas.
- After data cleaning, out of approximately 40,000 data points, was left with ~8400 unique entries.

**What is Success?**

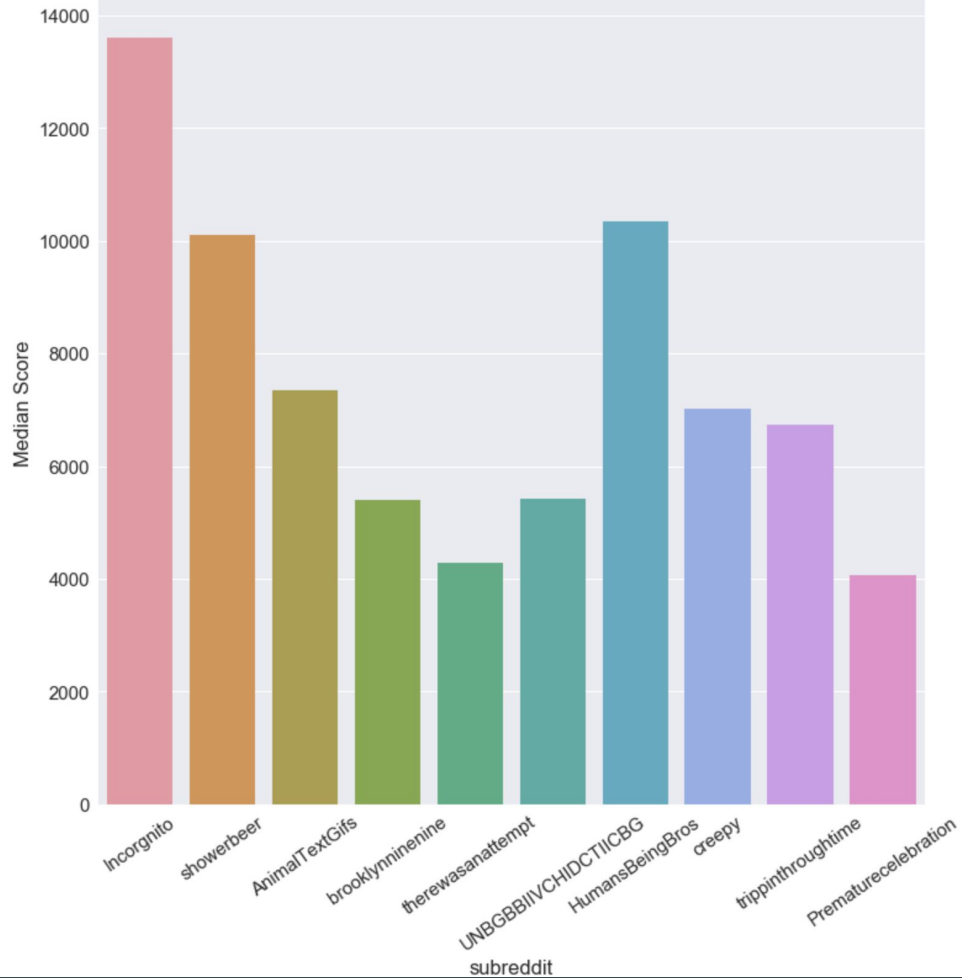
Ten subreddits with the Highest Median Score



The Ten Subreddits with the Highest Mean Score

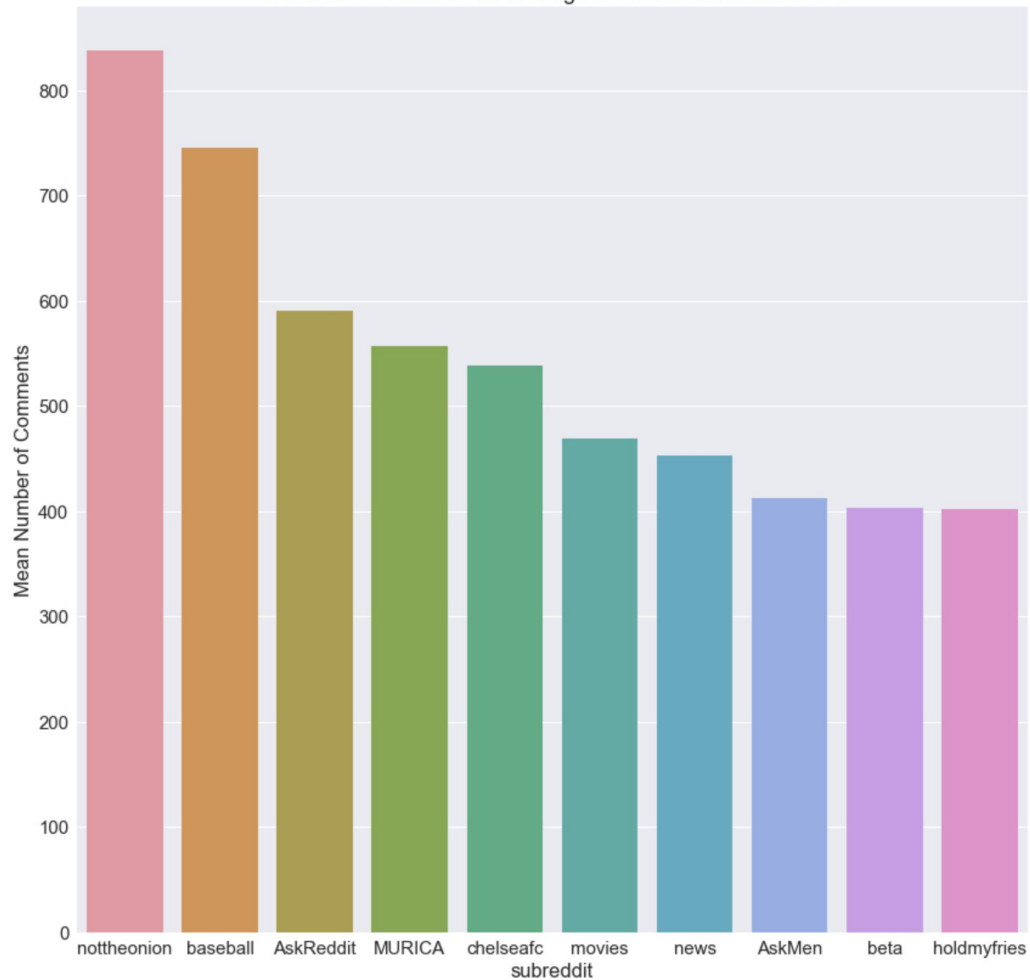


Ten subreddits with the Highest Median Comment Counts





The Ten Subreddits with the Highest Mean Comment Count



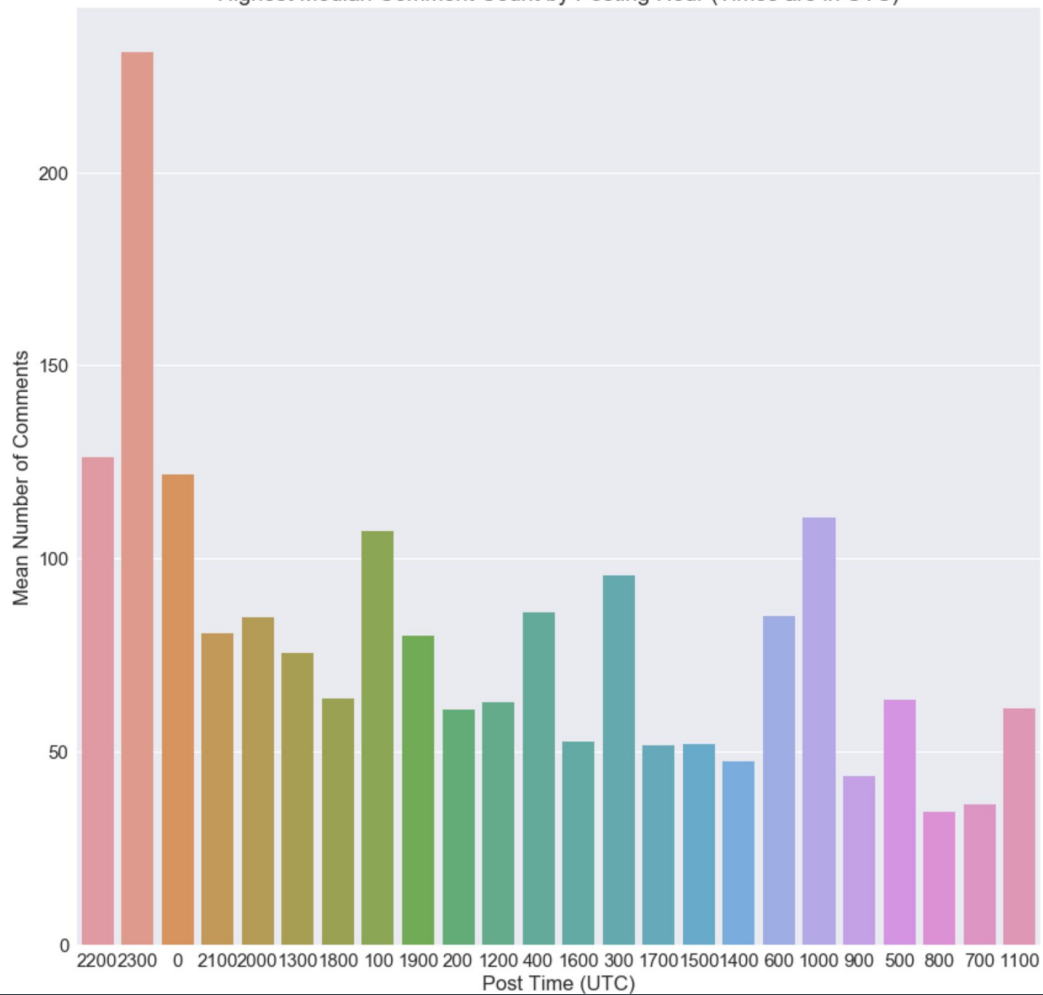
# Takeaways

- Certain subreddits (r/freebies, r/nottheonion) have consistently high scores,
- Posts here would have a good chance of getting good karma.
- In general, high scores on posts go hand in hand with high comment count.

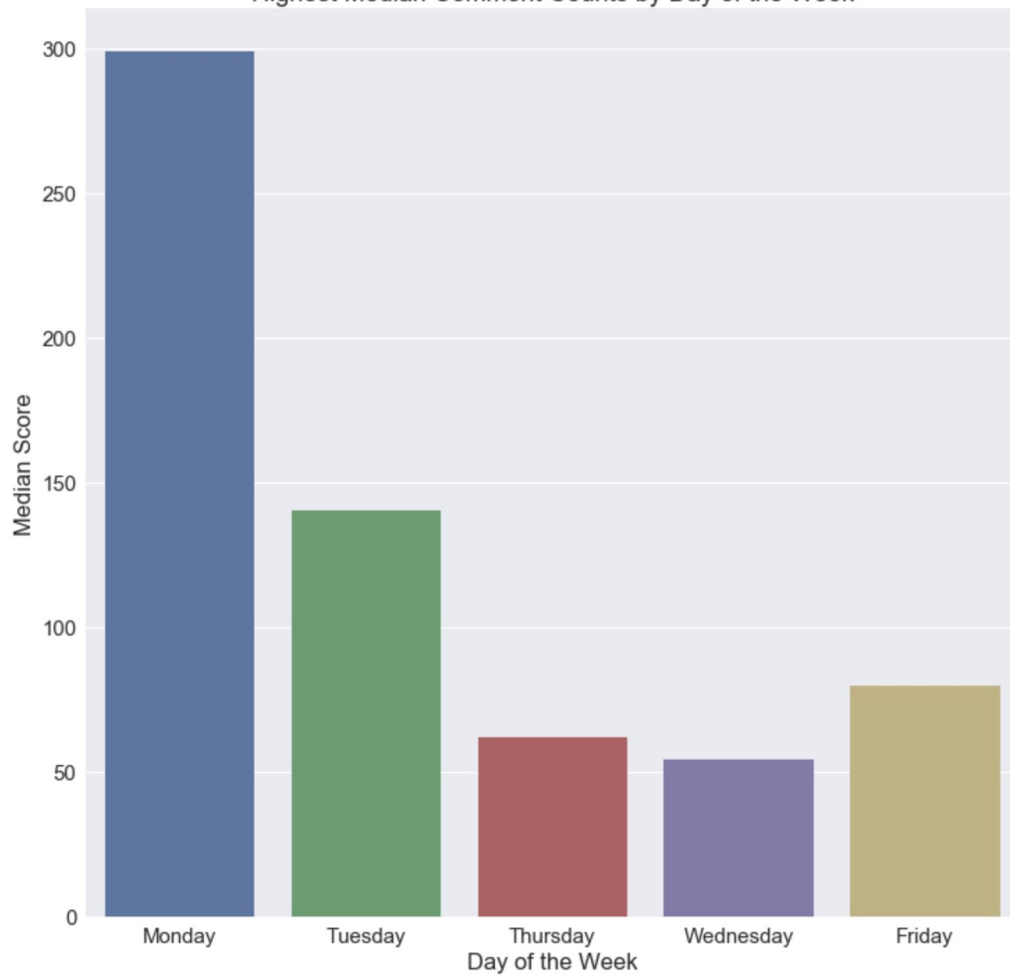
# Modeling Results

- Dataset was analyzed based upon percentile
- Post score and comment count were categorized based upon percentile.
- Subreddit was highly influential in determining if the post's comment count and score would land the post in the 95th percentile of posts.

Highest Median Comment Count by Posting Hour (Times are in UTC)



Highest Median Comment Counts by Day of the Week



# Potential Issues

- Scraped period was over a very short timeframe; more time (month(s)) would be ideal
- Post score is not always accurate
- Site is extremely fluid, changing on a day by day, sometimes hour by hour basis.
- Heavily influenced by current events

# Future Work

# Qualitative Data



