

This presentation may contain triggering and/or sensitive material.

The project analyzes user-submitted weight change posts on reddit.com. The data encapsulates such situations as users undergoing bariatric surgery to users recovering from severe anorexia.

No pictures are included, but please be aware of the sensitive topic at hand.



PYTHON AND PROGRESS PICS

August 27, 2020

TABLE OF CONTENTS

01

CONTEXT & HYPOTHESIS

Background on the dataset and
what I set out to discover

02

METHODOLOGY

Description of my code

03

RESULTS

What I discovered

04

NEXT STEPS

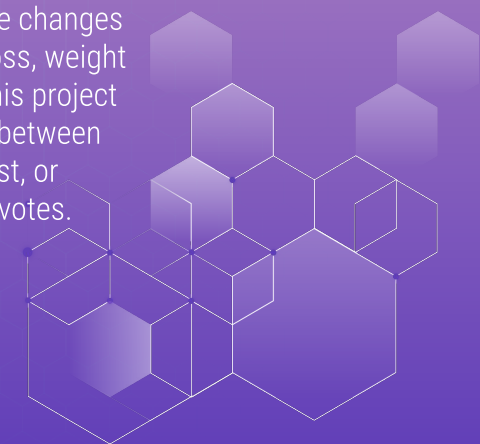
Ideas for taking this
analysis to the next level



CONTEXT

Reddit is an online social platform divided into "subreddits", which are smaller communities based around a common interest. Users can submit posts which are then "upvoted" or "downvoted" by other members of the community. Although community guidelines explicitly state the contrary, "upvoting" is often used to indicate approval or popularity of a post.

The r/ProgressPics subreddit is a community that is meant to share changes in physical appearance: weight loss, weight gain, body recomposition, etc. This project aims to see if there's correlation between age, gender, amount of weight lost, or number of comments against upvotes.





HYPOTHESIS

GOAL

This project aims to discover what variables lead to a higher upvote score on the r/ProgressPics subreddit. The variables under review include: age, gender, amount of weight change, and starting BMI, and number of comments.

HYPOTHESIS

I hypothesize that amount of weight change will be strongly correlated with number of upvotes. A larger absolute value will correspond with a higher number of upvotes.

METHODOLOGY

CREATE A DATAFRAME

Combine all necessary columns into one dataframe, ensure all data types are correct for analysis, and remove any errors or nonstandard formats that create errors.

INTERPRET

Confirm or disprove the original hypothesis.

CONNECT TO THE REDDIT API

Use PRAW (The Python Reddit API Wrapper) to connect to the Reddit API and pull the top 1,000 submissions out of the Progress Pics subreddit.

VISUALIZE AND ANALYZE

Create visuals and any other analyses that can help interpret the results



RESULTS

1

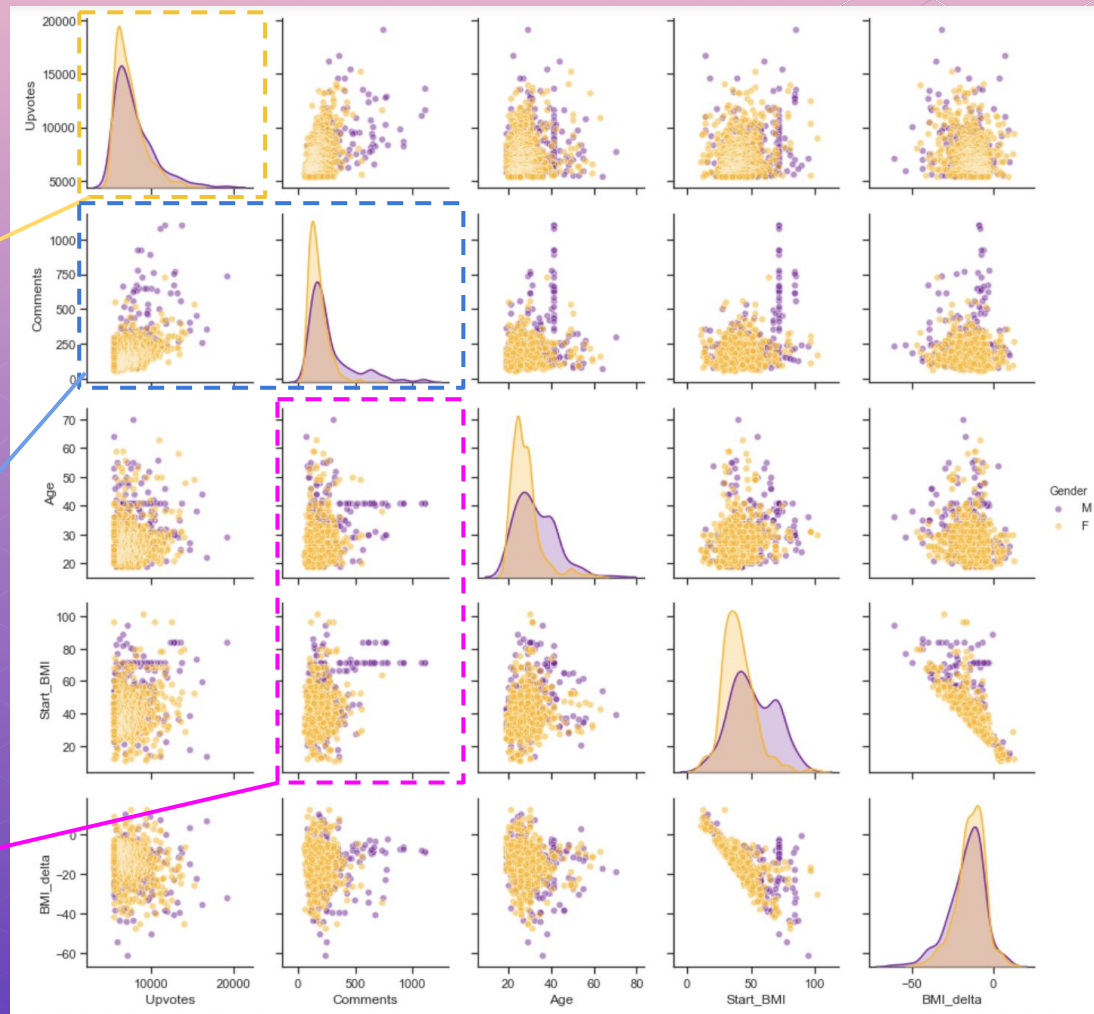
Women's posts are upvoted more frequently (yellow peak), but men's posts have more outliers where they receive more upvotes (purple tail).

2

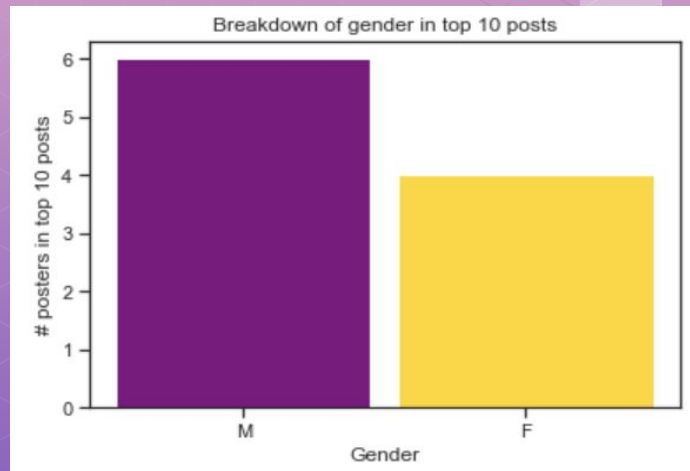
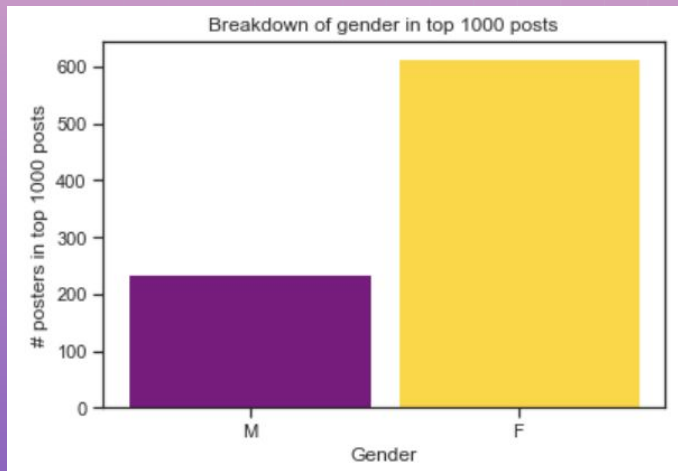
Women's posts are commented on more frequently (yellow peak shows about 1500 women receiving ~300 comments), but men's posts receive more comments (purple tail shows a handful of men receiving 1200+ comments; upper area mostly populated by purple dots.)

3

Long lines of repeated dots probably indicate a series of posts by the same poster. Individuals post update and progress shots of their journey at regular intervals.



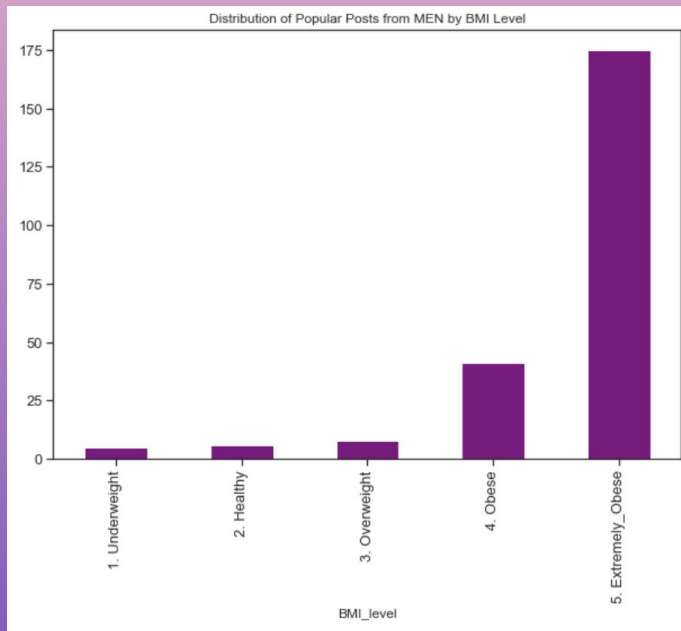
RESULTS



4

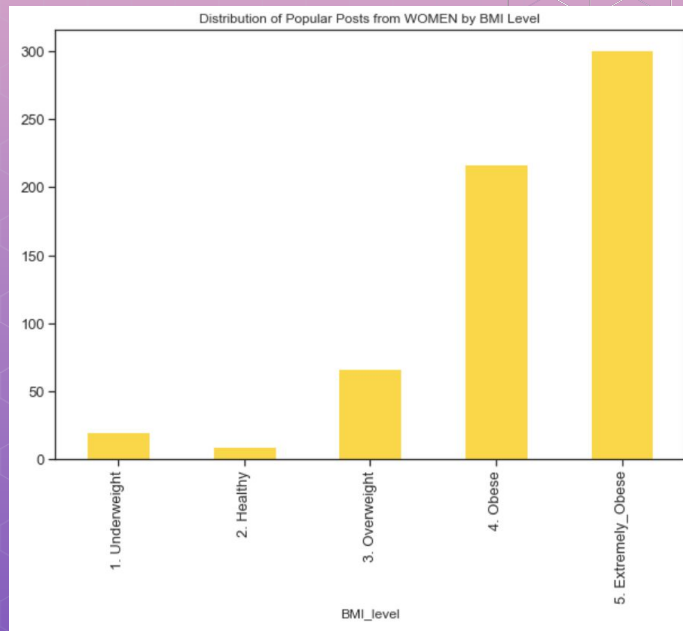
As shown on the previous slide, though more women than men are represented in the top 1000 posts (left bar graph), the top 10 submissions include more men (right bar graph).

RESULTS



5

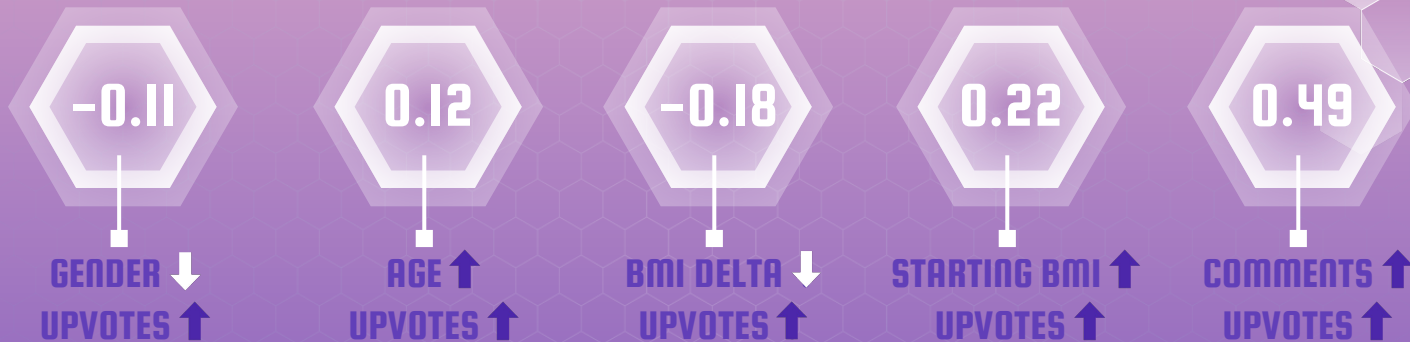
For men, there were about equal numbers of posts in the Underweight, Healthy, and Overweight categories. This can be interpreted to mean that male bodies were mainly upvoted when they were extremely large.



6

Interestingly, the pattern of upvotes for women was different: a wider variety of bodies were upvoted to the top of the subreddit. Moreover, the number of upvotes posts increase almost exponentially with higher categories.

RESULTS



7

Interestingly, there was a negative correlation between being a woman and amount of upvotes. This corresponds with the previous slide which showed that, though there were more women in the top 1000 posts, there were more men in the highest level of top posts.

8

Paradoxically, the negative correlation between BMI Delta and upvotes indicates that smaller weight changes have higher upvotes. This could be due to the fact that "series-posters" show minimal changes but develop a following because of their regular updates.

9

The strongest correlation was between comments and upvotes. Since both variables are proxies for engagement, this makes sense.

NEXT STEPS



Inspect the full dataset. 3x as many women comprise the top 1000, but perhaps that's because 10x as many women post overall. Further investigation is necessary.



Compare OP (Original Poster) comment count and upvotes. Are posters "rewarded" for taking part in the conversation?



Investigate "series-posters" further. What factors affect the number of upvotes over time for a single individual?



Convert metric units to imperial and include those submissions in the analysis.



THANK YOU