



**CAPSTONE PROJECT**

# **Does Reddit Know Who You Are?**

A look into User Personality and Behaviour  
Prediction on Reddit

# ABOUT REDDIT

- "Front Page of the Internet"
- 6th largest site in US/4th in Canada\*
- A massive collection of forums where you can share/post content and curate your own feed.
- Frontpage content based on user upvotes/downvotes
- Most Famous For: AMAs, Animal gifs...



# ABOUT MBTI

## MYERS-BRIGGS TYPE INDICATORS

Personality Classifiers based on Carl Jung's study. It classifies people based on the following 4 axis:

**Introversion (I) – Extroversion (E)**

**Intuition (N) – Sensing (S)**

**Thinking (T) – Feeling (F)**

**Judging (J) – Perceiving (P)**

That's 16 different personality classifiers.



# BUSINESS QUESTIONS

## WHAT IS THE STORY?

### KNOW YOUR AUDIENCE?

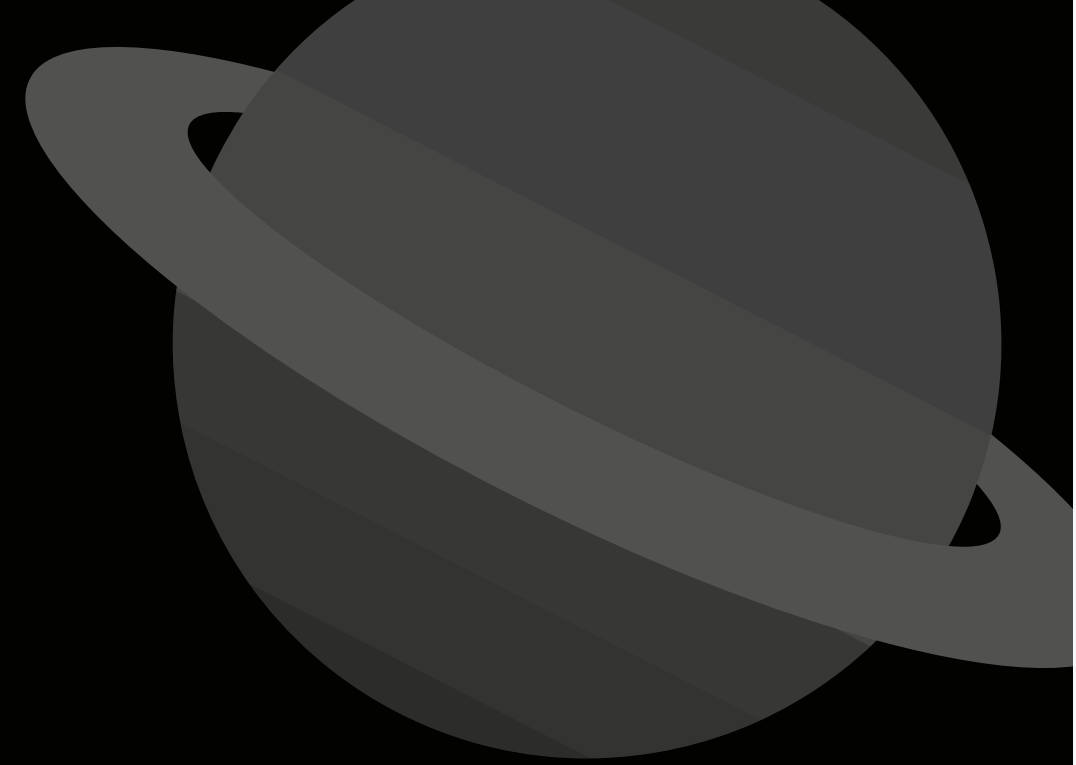
Can we predict the users' personality based on the comments they made?

### WHO ARE REDDIT USERS?

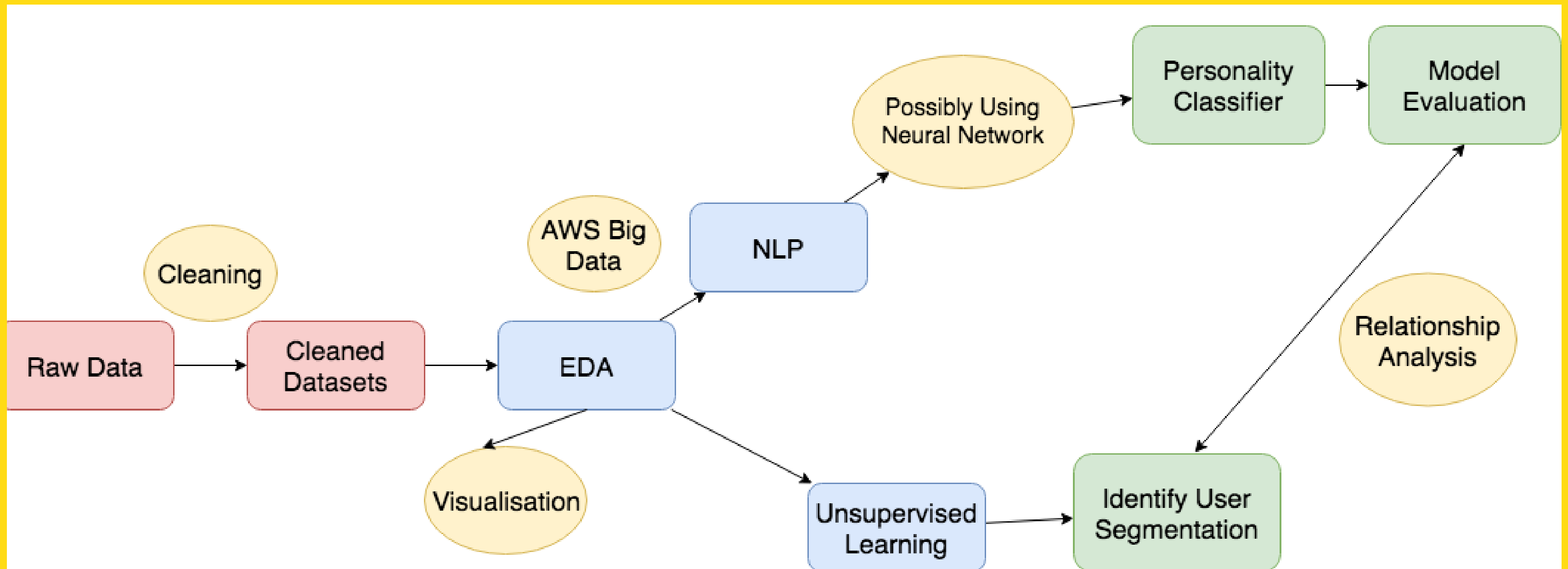
Can we build personas from the personality types and clusters?

### WHAT DO THEY LIKE?

Is there any obvious user segmentation based on the comments they posted and the subreddits?

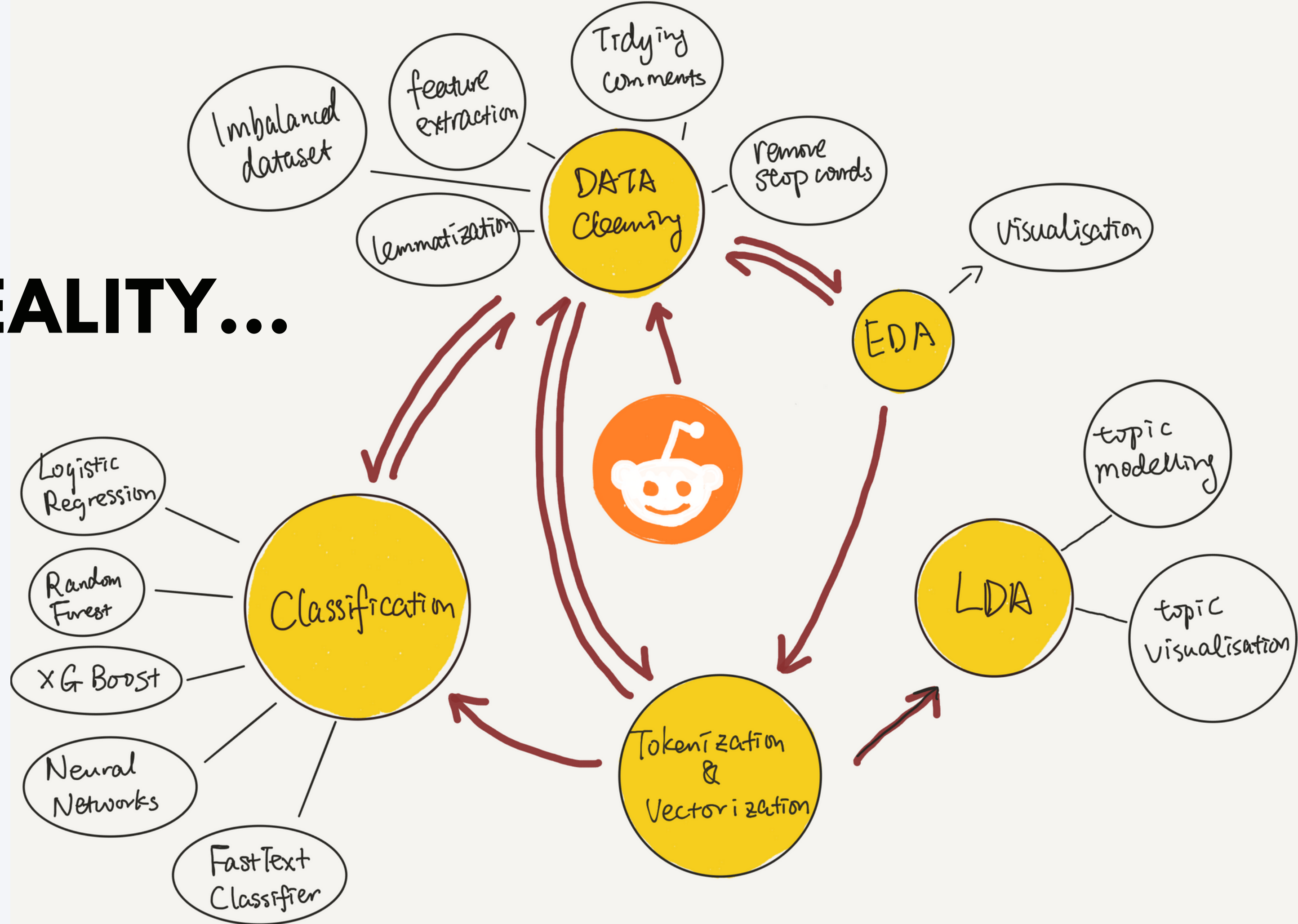


# PLANNED ROADMAP



**It looks straightforward,  
but....**

# IN REALITY...





**CLASSIFYING PERSONALITY BY  
SOCIAL MEDIA COMMENTS IS  
DIFFICULT...**

**HOWEVER, WE HAVE SOME  
GREAT TOOLS!**

**TOPIC MODELLING TO THE RESCUE!**



# LIMITATIONS

- Authority of MBTI (wildly applied by still pseudoscience)
- Imbalance of the data - also could be what the active Internet commenters are like (graph on the right)
- More supplementary datasets such as gender and age could help with clustering

