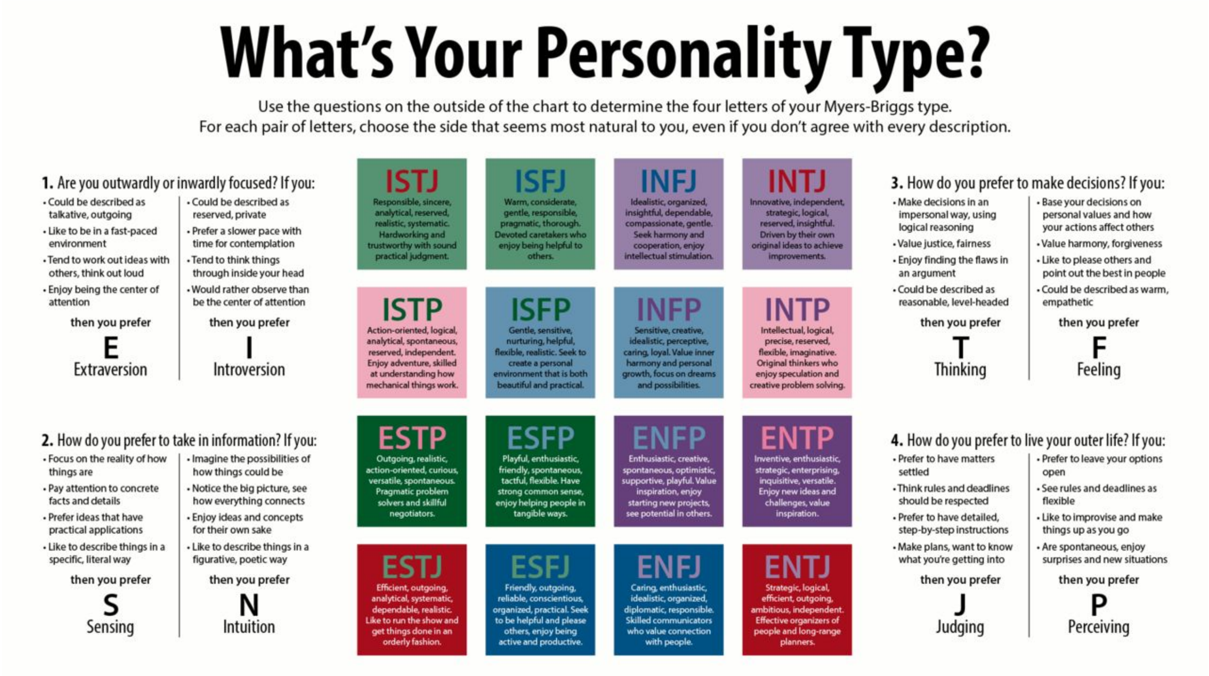
Project Outline

Determining MBTI’s personality type based on user's posts.

- **Explanation:** MBTI is a personality test indicator developed by Katharine Cook Briggs and Isabel Briggs Myers. It is based on the conceptual theory proposed by Swiss psychiatrist Carl Jung. It categorizes people into 16 personality types. My idea is to take the MBTI Kaggle dataset that classifies +8000 users from the website Personality Cafe into each of these 16 types and train a model to see if it is possible to determine personality types through written communication.

- **Problem Identification:** Uploaded on GitHub the problem identification slide which contains the following points:
  - Problem Statement
  - Context
  - Criteria for Success
  - Scope of solution space
  - Constraints within the solution space
  - Stakeholders to provide key insight
  - Key data sources

- **Framework:** The following image developed by Jake Beech provides a summary of the 16 personality types and the 4 axes.



What’s Your Personality Type?

Use the questions on the outside of the chart to determine the four letters of your Myers-Briggs type.
For each pair of letters, choose the side that seems most natural to you, even if you don’t agree with every description.

1. Are you outwardly or inwardly focused? If you:

- Could be described as talkative, outgoing
- Like to be in a fast-paced environment
- Tend to work out ideas with others, think out loud
- Enjoy being the center of attention

then you prefer

E
Extraversion

- Could be described as reserved, private
- Prefer a slower pace with time for contemplation
- Tend to think things through inside your head
- Would rather observe than be the center of attention

then you prefer

I
Introversion

2. How do you prefer to take in information? If you:

- Focus on the reality of how things are
- Pay attention to concrete facts and details
- Prefer ideas that have practical applications
- Like to describe things in a specific, literal way

then you prefer

S
Sensing

- Imagine the possibilities of how things could be
- Notice the big picture, see how everything connects
- Enjoy ideas and concepts for their own sake
- Like to describe things in a figurative, poetic way

then you prefer

N
Intuition

ISTJ Responsible, sincere, analytical, reserved, realistic, systematic. Hardworking and trustworthy with sound practical judgment.

ISFJ Warm, considerate, gentle, responsible, pragmatic, thorough. Devoted caretakers who enjoy being helpful to others.

INFJ Idealistic, organized, insightful, dependable, compassionate, gentle. Seek harmony and cooperation, enjoy intellectual stimulation.

INTJ Innovative, independent, strategic, logical, reserved, insightful. Driven by their own original ideas to achieve improvements.

ISTP Action-oriented, logical, analytical, spontaneous, reserved, independent. Enjoy adventure, skilled at understanding how mechanical things work.

ISFP Gentle, sensitive, nurturing, helpful, flexible, realistic. Seek to create a personal environment that is both beautiful and practical.

INFP Sensitive, creative, idealistic, perceptive, caring, loyal. Value inner harmony and personal growth, focus on dreams and possibilities.

INTP Intellectual, logical, precise, reserved, flexible, imaginative. Original thinkers who enjoy speculation and creative problem solving.

ESTP Outgoing, realistic, action-oriented, curious, versatile, spontaneous. Pragmatic problem solvers and skillful negotiators.

ESFP Playful, enthusiastic, friendly, spontaneous, tactful, flexible. Have strong common sense, enjoy helping people in tangible ways.

ENFP Enthusiastic, creative, spontaneous, optimistic, supportive, playful. Value inspiration, enjoy starting new projects, see potential in others.

ENTP Inventive, enthusiastic, strategic, enterprising, inquisitive, versatile. Enjoy new ideas and challenges, value inspiration.

ESTJ Efficient, outgoing, analytical, systematic, dependable, realistic. Like to run the show and get things done in an orderly fashion.

ESFJ Friendly, outgoing, reliable, conscientious, organized, practical. Seek to be helpful and please others, enjoy being active and productive.

ENFJ Caring, enthusiastic, idealistic, organized, diplomatic, responsible. Skilled communicators who value connection with people.

ENTJ Strategic, logical, efficient, outgoing, ambitious, independent. Effective organizers of people and long-range planners.

3. How do you prefer to make decisions? If you:

- Make decisions in an impersonal way, using logical reasoning
- Value justice, fairness
- Enjoy finding the flaws in an argument
- Could be described as reasonable, level-headed

then you prefer

T
Thinking

- Base your decisions on personal values and how your actions affect others
- Value harmony, forgiveness
- Like to please others and point out the best in people
- Could be described as warm, empathetic

then you prefer

F
Feeling

4. How do you prefer to live your outer life? If you:

- Prefer to have matters settled
- Think rules and deadlines should be respected
- Prefer to have detailed, step-by-step instructions
- Make plans, want to know what you're getting into

then you prefer

J
Judging

- Prefer to leave your options open
- See rules and deadlines as flexible
- Like to improvise and make things up as you go
- Are spontaneous, enjoy surprises and new situations

then you prefer

P
Perceiving

- **Expected Process:** I will follow the Data Science Method proposed by Aiden V. Johnson although later in the process we might need to make adjustments with the newly found information. The process goes as follows:
    - **Problem Identification:** As aforementioned, this document provides the starting point where I explain my initial hypothesis.
    - **Data Collection, Organization, and Definitions:** This section is very simple within this project since the data is readily available and complete. I will take the opportunity to create new features from the available data.
    - **Exploratory Data Analysis:** I will check for correlations to determine if there are variables we can drop, also to see trends and patterns that might give us more ideas for feature creation.
    - **Pre-processing and Training Data Development:** There are many pre-processing options at our disposal, like separating the posts, doing lemmatization, removing symbols and stopwords, doing sentiments analysis, etc. At the moment we do not know what is the best process in terms of what pre-processing steps will yield the best results.
    - **Model Creation:** With the newly generated dataset I will try to classify each user into a personality type (target column) by training classification models. I am currently not versed in this field but I imagine I use things like Support Vector Machines, K-Nearest Neighbors, Random Forest Classifier, etc. The process will include:
        - Fit Models with Training Data Set.
        - Review Model Outcomes - Iterate over additional models as needed.
        - Identify the Final Model
        - Apply the Model to the Complete Data Set
    - **Documentation**
        - Review the Results — Share your findings
        - Finalize Code and Documentation

- **Further analysis**: Maybe it could be interesting to do some follow-up analysis in this project. Some ideas that come into mind:
    - **Expand dataset:** Since the dataset is skewed, it might be interesting to scrape some more posts from the profiles that have less labeled data.
    - **Enneagram of Big 5 Models**: do a similar analysis with the enneagram or the Big 5 models (other personality test types) and see if there is a relationship between the tests. This has been a longstanding topic of discussion and could be interesting to shed some light here.
    - **LIWC:** The Linguistic Inquiry and Word Count (LIWC) is a text analysis tool that differentiates documents according to broad themes expressed in writing: psychological, relativity, and contextual themes. LIWC was created by Dr. James Pennebaker in order to examine relationships between language and personality. (See: M. C. Komisin) Using this technique we could extract some interesting insights as well.